



Université de Toulouse

MASTER 2 GEOMATIQUE

« **S**cience de l'**I**nformation **G**éoréférencée pour la **M**aitrise de l'environnement et l'**A**ménagement des territoires » (**SIGMA**)

<http://sigma.univ-toulouse.fr>

RAPPORT DE STAGE

Mise en place d'un atlas cartographique web de données agro- environnementales interopérable basé sur des technologies libres

FERRARI Hugo



Centre de coopération internationale en recherche agronomique pour le
développement

Tutrice du stage : Agnès TENDERO
Tuteur-enseignant : laurent JEGOU

Septembre 2015

Résumé

Pour améliorer l'efficacité des recherches, partager les données, faciliter leur accès par ses équipes, mieux maîtriser leur diffusion, communiquer ses résultats de recherches, le CIRAD Réunion souhaite regrouper l'ensemble de ses données géographiques sur un serveur web mettant en œuvre des solutions technologiques libres et respectant les règles d'interopérabilité. En l'état les ressources sont majoritairement stockées sur un disque réseau auquel l'accès est donné sur demande. De plus certains chercheurs produisent des données géographiques mais qui restent uniquement leur poste.

Il faudra donc à la fois partager les ressources mais aussi protéger la confidentialité des données présentées. L'équipe Artists (« Artists of Remote sensing Tools, Information systems, Simulation Techniques and Spatial analysis »), dont l'activité porte sur le traitement d'informations géographiques, s'est proposé pour porter ce projet et a entrepris sa maîtrise d'œuvre à travers un stage, rapporté dans le présent document. La solution d'une Infrastructure de Données Spatiales, ou IDS, correspondant à un système d'informations géographiques distribué sur Internet, s'adapte à la demande: A partir d'un modèle d'IDS existant appelé Geonode, qu'on a adapté au cahier des charges, le déploiement de l'atlas a été rendu possible dans les échéances prévues. Reste à convaincre la communauté d'utilisateurs potentielle d'adopter le produit. Le projet a donc inclus une phase de définition du besoin via un inventaire et un cahier des charges, avant la mise en œuvre concrète.

Mots-clés: catalogage, métadonnées, informations géographiques, logiciels libre

Abstract

With the purpose of improving research effectiveness, sharing datas, make their access easier for their team, Cirad in Reunion and Mayotte wants to gather all their geographical data sets on a web server, deploying open source or free softwares, and at the same time following interoperability rules. As things stand, resources are mostly stored in single network disks, and their access is given on demand. Furthermore, some reseachers create datasets, whichs only remain on their local machine. Resources will have to be shared respecting their privacy policy. team

Research team Artists (« Artists of Remote sensing Tools, Information systems, Simulation Techniques and Spatial analysis »), which works essentially on geographical datas, has planned to support this project and realized the operation through an internship, reported in this document. Spatial Data Infrastructure solutions fit the need, as they are Geographical Information System available on the Internet. From an IDS framework called Geonode, suited to the requirements specifications, atlas deployment has been possible before the deadline. Convincing potential user community to adopt the final software is the last part, after the internship. Project includes a first phase of requirement definition, through an inventory of data stored and requirement specification before the practical developpement.

Keywords: catalog, metadatas, geographical datas, free softwares

Table des matières

I.	Introduction.....	6
II.	Contexte.....	7
a.	Présentation du Cirad	7
	Statut, Historique et mission.....	7
	Organisation.....	7
b.	L'équipe de recherche Artists	8
	Dans l'organisation interne du Cirad.....	8
	Activité relative à l'île de la Réunion	8
	Activités et production	9
c.	Le constat qui motive le projet.....	10
III.	Etat des lieux	11
a.	Attendu et cahier des charges	11
	Spécifications.....	11
	Types d'utilisateurs	12
	Cas d'utilisations	12
	Contraintes	13
b.	Les données et métadonnées.....	14
	Dispositif matériel.....	14
	Inventaire des données	14
	Métadonnées	15
	Les applications existantes	16
	L'infrastructure de données spatiale.....	16
	Installation sur serveur (allocation ressources).....	18
	Echec avec easySDI , passage à Geonode	19
IV.	Développement	20
a.	Présentation de Geonode.....	20
	Fonctionnalités	20
	Les briques logicielles fondamentales	21
	Django.....	21
	Pycsw	22
	Geoserver	22
	les base de données du système	22
	Reproduction des scénarios d'utilisation évoqués dans le cahier des charges	22
	les permissions dans Geonode	24

b.	Adaptation de l'application aux nécessités du projet	25
	Outils de développement et d'administration système utilisés	25
	Inventaire des données par la gestion des permissions.....	26
	Gestion d'un groupe de contributeurs.....	26
	Tri des données selon les permissions	26
	Export métadonnées format HTML	27
	Authentification par le serveur de l'annuaire LDAP du Cirad.....	27
	Adaptation de la page d'accueil et mise en forme générale	28
	Modifications non-terminées.....	29
c.	Complétion du catalogue	29
	Reprojection préalable	29
	Données vectorielles	29
	Données Raster.....	30
	Complétion des métadonnées	30
V.	Résultat.....	31
a.	Bilan quantitatif.....	31
	En nombre de ressources.....	31
	Zones géographiques	31
	Producteur de données	31
	Volume des données.....	31
b.	Rapporté au cahier des charges.....	32
c.	Rapporté aux améliorations souhaitées.....	32
	La proposition de moissonnage des catalogues externes.....	33
d.	Bilan en termes de gestion de projet	33
VI.	Conclusion	35
VII.	Annexes.....	36
a.	Volumes de données sur le serveur Bebour: tableaux	36
b.	Schéma simplifié UML de la base de données "Geonode"	37
c.	Un Cas d'utilisation de l'atlas à Rodrigues.....	38
d.	Planning de Gantt, réalisé en début de projet (15 mars 2015).....	40
	Gantt Graphique	41
e.	Acronyme et glossaire.....	42
	Acronymes.....	42
	Glossaire et définitions.....	43
f.	remerciement	44

I. INTRODUCTION

Dans la recherche agronomique, l'information géographique est souvent indispensable : l'étude des relations entre les plantes cultivées, le sol, le climat et les techniques de culture, dont les principes régissent la pratique de l'agriculture, requiert l'usage de données spatialisées. De plus, la cartographie y revêt un rôle primordial, utilisée à la fois au cours des travaux de recherche, puis pour la diffusion des connaissances. Les technologies informatiques en ont décuplé les capacités: l'acquisition des données, leur traitement et leur représentation cartographique sont assistés par ordinateur.

En revanche, la diffusion des données reste un sujet épineux malgré l'apport de l'informatique; d'abord pour les aspects techniques, car il y a mille manières de produire l'information, et en l'absence de coordination, le milieu de la recherche pourrait s'apparenter à une tour de Babel, où chacun produit ses données à sa manière sans chercher à les communiquer ultérieurement. Ensuite, apparaissent les problèmes de propriété et de confidentialité, les travaux de recherche pouvant faire l'objet de rétention.

La direction régionale du Cirad (Centre de coopération Internationale en Recherche Agronomique pour le Développement) à La Réunion, en tant qu'établissement de recherche, s'est déjà munie de la plupart des outils numériques de production de données géographiques. Cependant, la diffusion des données est toujours limitée: conservées de manière éparse, elle n'est pas valorisée et généralement les membres de l'institution ne peuvent y accéder simplement, ou même simplement savoir qu'elle existe.

L'équipe Artists (pour « Artists of Remote sensing Tools, Information systems, Simulation Techniques and Spatial analysis »), œuvrant au Cirad à La Réunion dans le domaine du traitement de l'information géographique et d'images satellitaires, se propose de faciliter la diffusion et l'accès à la donnée en réalisant un atlas des données qu'elle détient, plus celles détenues par les autres équipes de recherche ou partenaires institutionnels de l'établissement. Consultable sur Internet, premier réseau de diffusion de l'information numérique, il utilisera des technologies libres de droits.

Il est question, dans le présent document, du stage dédié à la concrétisation de ce projet. Nous préciserons tout d'abord le contexte général: le Cirad et le territoire Réunionnais ; nous reviendrons sur le cahier des charges ainsi que l'inventaire des données existantes qui mènent aux propositions de mise en œuvre. La majeure partie concerne la mise en œuvre elle-même : nous justifierons les choix technologiques en détaillant le fonctionnement du système développé. Enfin nous rapporterons le résultat obtenu à la commande initiale, aux améliorations possibles et établirons un bilan de la gestion du projet.

II. CONTEXTE

a. PRÉSENTATION DU CIRAD

STATUT, HISTORIQUE ET MISSION

Le Centre de coopération International en Recherche Agronomique pour le Développement, Cirad, placé sous la double tutelle du ministère de l'enseignement supérieur et de la recherche et du ministère des affaires étrangères et européennes, est un établissement public à caractère industriel et commercial (EPIC). A ce titre, en plus des travaux de recherche, les équipes qui le composent peuvent proposer des services et produits à leurs partenaires institutionnels. Au service du développement des pays du Sud et de l'Outre-Mer français le Cirad mène ses recherches dans le domaine des sciences du vivant, des sciences humaines et des sciences de l'ingénieur appliquées à l'agriculture et à l'alimentation, à la gestion des ressources naturelles et aux sociétés¹.

A l'origine de sa fondation en 1984, se trouvent les neufs instituts de recherche agricole tropicale, pour la plupart datant des années quarante. Ces derniers se sont rassemblés en 1958 au sein d'un comité de liaison des organismes de recherches agricoles spécialisés outre-mer, puis réunis en 1970 pour former le groupement d'étude et de recherche pour le développement de l'agronomie tropicale, dont est issu le Cirad.

L'institution privilégie la recherche scientifique par des expérimentations, des actions de formation, d'information et d'innovations, et enfin d'expertises. Pour mener à bien ses missions, le Cirad emploie 1750 agents, dont 800 chercheurs, et disposait en 2014 d'un budget de 197 millions d'euros.

ORGANISATION

Le Cirad est implanté en France métropolitaine et dans l'outre-mer mais aussi dans de nombreuses régions du monde, ce qui lui permet de travailler en collaboration avec plus de 90 pays.



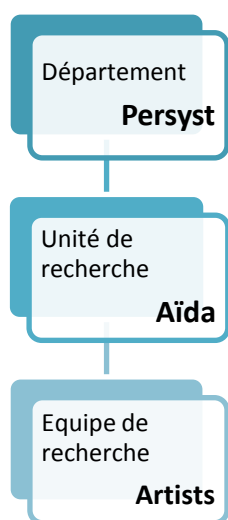
Figure 1 Les directions régionales du Cirad dans le monde. En jaune la position du stagiaire. Source Cirad.fr

¹ Source : Cirad.fr

Indépendamment de la localisation des pôles de recherche, les agents sont répartis entre 3 départements scientifiques et la Direction Générale à Paris, qui sont composés d'unités de recherches (UR), 37 au total. Dans chaque UR, on trouve plusieurs équipes de recherche, dont Artists, à l'origine de ce stage. Nous allons décrire succinctement les activités de cette équipe, situer sa place au sein du Cirad d'abord, puis par rapport à l'activité agricole Réunionnaise ensuite.

b. L'ÉQUIPE DE RECHERCHE ARTISTS

DANS L'ORGANISATION INTERNE DU CIRAD



Le département Performance des systèmes de production et de transformation tropicaux (Persyst), porte ses recherches sur l'élaboration et la gestion intégrée de la qualité des productions agricoles et des produits agroalimentaires, et conduit des études sur les productions tropicales (agriculture familiale et production de rente) à l'échelle de la parcelle, de l'exploitation et de la petite entreprise de transformation. Aïda (Agroécologie et Intensification Durable des Cultures Anuelles), est une de unités de recherche du département Persyst. Elle regroupe 90 personnes dont 55 chercheurs, et séparée en 5 équipes de recherche, animées par la volonté collective de mettre en œuvre des démarches interdisciplinaires au service de l'intensification écologique. Parmi ces équipes se trouve Artists.

ACTIVITÉ RELATIVE À L'ÎLE DE LA RÉUNION

Le stage s'est déroulé au sein de la direction régionale du Cirad à La Réunion et à Mayotte. Les 180 salariés, dont 55 chercheurs, sont pour la majorité situés sur l'île de la Réunion². Cette direction existe depuis 1962 sur ce qui est actuellement un Département et une Région d'Outre Mer. De quoi est fait le territoire Réunionnais et en quoi les travaux de l'équipe Artists intéressent les partenaires locaux ? Pour appréhender l'activité de l'équipe Artists, il faut rappeler quelques précisions sur le territoire Réunionnais.

Au premier janvier 2009, on comptait sur l'île de La Réunion 800 000 habitants (INSEE), pour une densité d'environ 282 hab/km². Le relief escarpé de l'île fait que l'essentiel de la population vit le long des côtes ; là où se concentre également le plus gros de l'activité économique. L'agriculture participe à 2% du PIB Réunionnais. Le relief montagneux et accidenté de l'île explique que la SAU (Surface Agricole Utilisée) ne représente qu'un cinquième de sa superficie (près de la moitié en métropole). La canne à sucre occupe 60% de la surface agricole utilisée³. Ce qui fait de la Réunion le premier producteur de sucre de canne à destination de l'Union Européenne. Ainsi l'activité cannière présente une part importante de l'agriculture réunionnaise. L'équipe Artists réalise entre autres des

² Source : reunion-mayotte.cirad.fr

³ Source reunion-mayotte.cirad.fr

prévisions de récoltes par une analyse d'images satellitaires doublée d'un modèle de prévision, à la demande de la filière canne.



Saint Pierre,
Île de La réunion
*Ce cliché reflète l'occupation
des sols de l'île, avec une bande
côtière occupée par des
logements et dans les terres un
parcellaire très morcelé.*
Source
<http://www.ipreunion.com>

ACTIVITÉS ET PRODUCTION

L'équipe Artists est animée par Pierre Todoroff, chercheur en modélisation et télédétection. Ses membres permanents sont Agnès Tendero, ingénieure en informatique et responsable des projets SI et outils d'aide à la décision, Mickael Mézino, ingénieur en informatique et administrateur des systèmes et des bases de données, Lionel Le MEZO, ingénieur en télédétection, Louis PAULIN, technicien et écophysiologiste, Raymond Nativel, technicien et métrologue et enfin Dany Deurveilher, technicien et machiniste/télédéacteur

L'équipe a donc comme première vocation la production de travaux de recherche à partir de la télédétection sur la filière agricole à la Réunion. Mais elle fournit notamment des services auprès des agriculteurs, en expertise et conseil notamment. Ses deux axes de recherche sont :

- La gestion de l'information spatiale. Disposer des données d'entrée pour alimenter les modèles et outils d'aide à la décision, est le premier maillon de la chaîne de traitement de l'information. Une des productions majeure est la plateforme web [Margouill@](#): Modélisation pour l'Agriculture Réunionnaise par Géoréférencement et Outils et Logiciels Libres sur le Net. L'aspect innovant de ce nouveau système réside notamment dans le couplage entre modèles agronomiques et Système d'Information Géographique (SIG) permettant ainsi de représenter simultanément les phénomènes spatiaux et temporels, interactions déterminantes de la production végétale à l'échelle régionale.
- La modélisation et la création d'outil d'aide à la décision. L'équipe Artists prolonge ses travaux de développement de système d'information par des analyses des données recueillies aboutissant à des modèles ou représentations de ces données, et la connexion de modèles existants ou de nouveaux modèles comme modules complémentaires aux bases de données. Les techniques de télédétection occupent une large part des développements méthodologiques de cet axe.

L'imagerie satellitaire est une source d'information particulièrement adaptée pour recueillir des informations exhaustives sur de grandes régions ⁴.

C'est le couplage entre les modèles et les données contenues dans les systèmes d'information qui rend possible la valorisation des modèles sous forme d'outils spatialisés. Ces outils sont ensuite fournis sous forme de prestations aux partenaires institutionnels.

C. LE CONSTAT QUI MOTIVE LE PROJET

Au sein du Cirad à la Réunion, l'équipe Artists remplit une autre tâche : fournir aux autres équipes de recherche des données géographiques dont elles ont besoin. Seulement, les agents ne disposent pas toujours des capacités et des moyens techniques d'Artists en termes de capacité de stockage. En général on s'adresse donc à l'équipe pour récupérer la donnée. Il n'existe pas d'inventaire qui permette aux membres du Cirad de voir ce qui existe ni d'outil de recherche adapté à leur besoin. Les agents méconnaissent donc souvent l'existence des données, et c'est suite à une demande verbale (par exemple: quelles sont les données d'occupation du sol sur le territoire de Madagascar ?) que s'effectue la transaction. Qui plus est les informations relatives à ces données (date de création, propriétaire, qualité) sont parfois indisponibles et rendent difficile leur utilisation rigoureuse dans le cadre d'une recherche scientifique. Devant ce besoin collectif et après un recensement des chercheurs et équipes intéressés, il a été convenu que plusieurs UR financeraient le stage, dans la mesure où le projet pouvait leur bénéficier directement. Il s'agira donc de réaliser un outil qui permette aux utilisateurs concernés d'accéder à la donnée géographique.

A moyen terme, l'on cherche à valoriser et à mutualiser les données réalisées dans le cadre des travaux de recherche. En effet, un certain nombre de jeux de données sont uniquement sur les postes des agents les ayant réalisées. Avec l'atlas, ces données pourraient être rendues visibles, tout en préservant leurs contraintes d'accès.

⁴ Source: ur-aida.cirad.fr

III. ETAT DES LIEUX

Nous allons donc présenter l'état des lieux qui précède la mise en œuvre du projet. Nous répondrons principalement dans cette partie à trois questions:

- Quel est l'attendu précis du projet ?
- Comment caractériser les données entreposées sur serveur ?
- Quelles sont les solutions que l'on peut proposer ?

Nous abordons la situation comme elle se présente au début du stage: De quoi a-t-on besoin, quel est la donnée à notre disposition. De là nous établissons une série de propositions.

a. ATTENDU ET CAHIER DES CHARGES

En premier lieu il faut procéder à la définition d'un cahier des charges. Cette phase s'apparente à une maîtrise d'ouvrage, car l'on décrit avec les commanditaires directs (les membres de l'équipe Artists), en quoi va consister l'atlas. Il faut comprendre qu'au début du stage, aucune solution technologique n'est arrêtée, et que toutes les hypothèses de cas d'utilisation peuvent être envisagées. Ainsi, on part sans limitations induites par des choix technologiques préalables.

Qu'entend-on par Atlas Cartographique? La définition stricto sensu est un recueil de cartes géographiques, ordonné selon un critère spécifique, l'ordre alphabétique par exemple, ou la localisation. Il existe donc dans la définition de base l'idée d'inventaire, de liste. C'est le seul point commun que l'on peut trouver avec un atlas version papier. L'idée du projet est plus à rapprocher d'une cartothèque, dans la mesure où il s'agit de disposer de ressources, et de les extraire de leur ensemble de départ pour les utiliser. Seulement, la multiplicité des types de données géographiques rend le terme de cartes caduques. On parlera dans la suite du rapport de « ressource », de manière équivalente au terme « donnée ». De la même manière, un inventaire de données informatiques classées selon leurs métadonnées est désigné sous le terme de catalogue.

En plus de l'atlas lui-même, on attend du stagiaire qu'il rencontre les équipes de recherche qui seront parmi les contributeurs de l'atlas. Une fois l'atlas finalisé, une phase de formation devra être fournie aux futurs utilisateurs, membres du Cirad, et doit être assurée par le stagiaire, accompagné des membres de l'équipe.

On replace dans cette partie les attendus présentés en amont du projet : il ne s'agit donc pas de ce que l'application finale réalise, mais de ce qu'on en attendait avant le développement

SPECIFICATIONS

- L'environnement d'utilisation est un navigateur Web. Les technologies à utiliser doivent être des *logiciels libres*
- Le projet est destiné d'abord aux membres des équipes du Cirad à la Réunion. Ce sont les utilisateurs ciblés en priorité. Cependant, la plateforme doit aussi permettre à un utilisateur anonyme d'accéder à un inventaire des données détenues au Cirad à la Réunion
- Les données renseignées dans ce projet seront dans l'emprise de l'océan Indien, dans les zones d'étude des membres du Cirad à La Réunion (Madagascar - îles Mascareignes, Mayotte mais aussi

côte Est de l'Afrique, Comores, etc). Cela n'exclut pas l'ajout de données extra-océan Indien, mais donne une idée de l'emprise principale.

TYPES D'UTILISATEURS

On perçoit trois types d'utilisateurs:

- L'utilisateur qui consulte de données : il peut visualiser la liste des données et les récupérer sur son poste.
- Le contributeur qui en plus peut ajouter des nouvelles données.
- L'administrateur qui veillera à la qualité des ajouts en validant les contributions

Cette vision simplifiée permet d'anticiper des scénarios d'utilisation.

CAS D'UTILISATIONS

Les scénarios d'utilisations correspondent aux deux fonctionnalités majeures: "l'utilisateur fait une recherche", et "l'utilisateur ajoute une donnée" (on écarte pour l'instant "l'utilisateur valide la donnée"). Chacun de ces scénarios peut être interrompu à n'importe laquelle de ses étapes:

A Consultation de la donnée disponible

1. l'utilisateur fait une requête de sélection de type

- Texte, Selon des Critères
 - Nom de la donnée
 - Thématique : quel sont les grand thèmes que couvrent la donnée, et comment les regrouper ?
 - Auteur/Fournisseur/ producteur
 - format de la donnée
 - date de production
 - localisation
- Géographique, sur fond de carte: on imagine une carte interactive sur laquelle l'utilisateur peut zoomer et ainsi définir une emprise d'inclusion de sa recherche.

2. L'utilisateur visualise le résultat

- Sur un fond de plan (afficher l'emprise de chaque résultat de la recherche sur un fond de carte)
- Par une liste de résultats, avec un résumé pour chaque occurrence trouvée

3. L'utilisateur sélectionne une donnée

4. L'utilisateur visualise une donnée en particulier: accès à la métadonnée complète, ainsi qu'à une visualisation via une interface de Webmapping

5. L'utilisateur rapatrie la donnée sur son poste

B – ajout d'une nouvelle donnée

1. l'utilisateur renseigne les informations (métadonnées) sur une donnée
2. L'utilisateur ajoute la donnée au système.
3. Un administrateur valide la donnée ajoutée et la rend visible pour le cas d'utilisation A

Ces définitions triviales permettent de définir un cadre aux attendus de l'application: elle doit permettre à minima de dérouler ces scénarios.

CONTRAINTES

L'utilisation de la plateforme doit s'inscrire dans le cadre de contraintes.

- Confidentialité des données et sécurisation du serveur

Certaines données ne peuvent pas être rendues disponibles à tous les utilisateurs sans l'accord préalable des producteurs de la donnée ou d'autres personnes concernées par sa diffusion. Cela distingue plusieurs catégories de visiteurs pour consultation.

- Ressources matérielles

Si l'espace de stockage disque sera alloué par la direction des services informatiques (DSI) selon le besoin déterminé, les ressources en terme de capacité de calcul du serveur Web et de mémoire alloué sont des facteurs plus limitants. Il faudra veiller aux performances de l'application durant son développement

- Sécurisation de l'application

En déployant l'application sur un serveur Web on s'expose à des attaques externes, si l'application n'est pas bien protégée. On pense notamment aux attaques informatiques ayant pour but de récupérer une donnée confidentielle, des identifiants de connexion ou simplement de dégrader le système informatique.

- Interopérabilité avec les systèmes existants

Par Interopérabilité, on entend: la capacité que possède le système, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre. Ainsi l'application doit pouvoir fonctionner avec les systèmes de données géographiques déjà utilisés au Cirad.. De plus certaines administrations ou associations ont déjà monté leurs applications, qu'il conviendra d'étudier. Ce qui découle de la première contrainte d'interopérabilité est la nécessité renseigner les métadonnées selon une norme existante: en effet il est plus simple d'échanger des informations si elles sont structurées de la même façon, selon une norme appliquée par tous.

La conformité à INSPIRE

La directive INSPIRE est issu de l'Union Européenne de 2007, et vise à assurer l'interopérabilité entre bases de données géographiques pour faciliter la diffusion, la disponibilité, l'utilisation et la réutilisation de l'information géographique. Pour y parvenir, il faut se plier à une série de contraintes lors du développement et de l'utilisation d'un système informatique. La question est: Le Cirad est-il obligé de se plier à cette directive et sinon, est-ce vraiment intéressant de l'appliquer tout de même?

D'après une note du labex CEBA⁵ (Centre d'Étude pour la biodiversité amazonienne) datant de juillet 2013, «les organismes de recherche ne sont pas concernés par les textes réglementaires relatifs à la liberté d'accès et la ré-utilisation des informations publiques et ne sont donc pas soumis à la directive Inspire». Les données que produit le Cirad servent des fins expérimentales ou de recherche, leur diffusion n'est pas soumise à une obligation légale. Cependant, certaines des données sont diffusées auprès des filières agricoles et des industriels comme produit finalisé. Par exemple la carte de suivi de la coupe peut être utilisée par les industriels comme outil d'aide à la planification de la

⁵ Source docs.gip-ecofor.org

campagne sucrière. La production de cartes d'occupation du sol agricole basé sur l'analyse d'images satellites est à destination des institutions publiques (Commune, Préfecture etc). Ce cadre de diffusion n'oblige cependant pas à respecter la directive INSPIRE. Ainsi, si l'on s'y conformait, ce serait uniquement par soucis d'interopérabilité avec les systèmes existants, et non pas par obligation légale.

b. LES DONNÉES ET MÉTADONNÉES

Nous allons décrire les données de deux manières différentes: comment elles sont stockées, puis leur description qualitative et quantitative au travers notamment d'un inventaire

DISPOSITIF MATÉRIEL

En l'état les données qui seront intégrées à l'atlas sont stockées de trois manières différentes:

- Un disque réseau du serveur de données géré par la direction des Services Informatiques, nommé **Bebour**. Les données concernent uniquement La Réunion et sont triées selon le producteur de la donnée (IGN, ONF, DAAF). Ce disque réseau est protégé: il faut un mot de passe pour y accéder.
- Dans une base de données géographique PostGIS, sur laquelle sont stockées les données servies par l'application web **Margouill@** décrite dans la précédente partie.
- Sur un serveur propre à l'équipe **Artists**, qui cette fois contient non seulement les données sur la Réunion en doublon sur **bebour**, mais aussi des données sur d'autres pays. Il est à noter que l'accès à ce serveur est exclusif aux membres d'Artists. Deuxième fait notable, il contient toutes les images satellitaires, et de ce fait sa capacité de stockage est la plus importante, avec 7,2 To disponibles.

INVENTAIRE DES DONNÉES

Description qualitative

Format et type

Généralement, on sépare en deux catégories principales la donnée géographique: la **représentation** vectorielle et la représentation raster. Présentons les plus importants formats correspondants à ces représentations, trouvées dans l'atlas:

- le format **Shapefile**, de type vectoriel
- le format image **Geotiff** et le format ECW (Enhanced Compressed Wavelet) sont de type raster. les données de type grille peuvent aussi être des matrices de valeur au format texte, qui peuvent être affichées
- Enfin, on trouve des cartes "finales" c'est à dire dans leur forme définitive avec légende et échelle intégrée, au format PDF ou jpeg. Ces documents n'ont pas de références spatiales, mais il serait pertinent de les faire figurer dans l'application. De plus, certains écrits, comme les notices d'utilisation ou les licences détaillées, permettraient à l'utilisateur de connaître le cadre d'utilisation des données de l'atlas.

Producteurs

Le producteur le plus référencé est sans surprise l'Institut Géographique National (IGN), qui est l'organisme de référence pour la production de données géographiques en France. Le Cirad à la Réunion dispose pour ces données d'une licence d'utilisation spécifique.

Les autres producteurs sont des organismes d'état comme: l'Office National des Forêts, la Direction Régionale de l'Eau et du Logement, le Parc Naturel, le Bureau de Recherche Géologique et Minière. Fait important, la plupart de ces producteurs ont eu pour obligation dans le cadre de la directive INSPIRE, de rendre leurs données publiques: on peut donc les retrouver sur Internet gratuitement, avec seulement quelques conditions d'utilisations et de diffusion. On notera la présence de données nominatives, c'est à dire associant aux entités géographiques des personnes physiques ou morales, dans le cadastre notamment. Ces données sont considérées comme sensibles et l'équipe Artists a souhaité restreindre leur diffusion aux membres de l'équipe seulement.

A Madagascar il existe le FTM, équivalent de l'IGN. L'équipe Artists s'est procuré une base de données géographiques sous forme de fichiers Shapefile, mais cette acquisition s'est fait dans le cadre d'un projet de recherche unique. Ainsi ces données seront elles aussi diffusées exclusivement auprès de l'équipe Artists.

Thématique

Les thématiques des données sont celles relatives à l'agronomie, et plus largement à la géographie.

Estimation quantitative

On peut établir une estimation des données selon les critères qualitatifs ([voir annexe 1](#)). Sur les 41,7 Go de données sur la Réunion par exemple:

- le Cirad produit finalement peu de données rapporté au total des données en présence. En effet, seuls 0,7 Go de données sont comptabilisés comme issus du Cirad. Ce chiffre laisse à penser que le Cirad produit peu, mais nous n'avons pas comptabilisé les données des autres équipes de recherche, qui n'ont pu être récupérées en amont du projet. En l'état, la plupart des données stockées viennent de l'extérieur.
- le format le plus largement représenté, en nombre de fichiers, est le format Enhanced Compressed Wavelet (ECW). Utilisé pour la compression d'image raster, il a l'avantage d'être réputé le plus performant par son taux de compression. L'inconvénient est qu'il n'est pas libre: il faut des logiciels propriétaires pour l'utiliser. Souvent, les solutions libres ne proposent pas le support de ce format. La solution pour contourner ce problème est détaillée dans [cette partie](#).
- Les images satellitaires utilisées pour les projets de télédétection, stockés sur le serveur Artists, occupent 2300 Go de données, c'est 43 fois plus que le reste (les données PostGIS et fichiers sur Bebour). Or, ces données sont exclusivement utilisées par les membres de l'équipe Artists, à des rares exceptions près. Compte tenu de leur volume, **elles ne seront pas mises à disposition sur l'atlas**, ou alors sous forme d'échantillon.

Finalement, les données en présence sont surtout des données de référence plus que des productions réalisées en interne. On trouvera donc des données en provenance des grands fournisseurs nationaux français.

MÉTADONNÉES

La métadonnée est le lot d'informations servant à décrire un jeu de données: résumé, date de production, édition, entre autres. Les métadonnées sont indispensables au fonctionnement de l'atlas, car c'est par leur biais que l'utilisateur va effectuer ses sélections. Or, si les grands producteurs nationaux comme l'IGN sont maintenant légalement contraints à produire des métadonnées, il n'en va pas de même pour tous les organismes; Pour les données du Cirad en particulier, il existe des

carences, qu'il faudra combler en interrogeant notamment les membres de l'équipe Artists qui les connaissent.

LES APPLICATIONS EXISTANTES

Nous parcourons ici les applications qui peuvent remplir partiellement les objectifs du projet attendu. Les solutions internes et externes sont observées d'un point de vue technique et pratique.

Le catalogue de métadonnées interne au Cirad

Le catalogue actuel est sous forme de fiches au format XML, et ne renseigne que les données présentes sur le disque Bebour. Ces métadonnées ne sont pas conformes à INSPIRE. Le catalogue a été assemblé pour la dernière fois en 2010, il n'est clairement pas à jour. Il est possible de parcourir les fiches dans un navigateur web, selon une arborescence dont le premier niveau est le producteur de données. Il n'est possible de filtrer que sur le titre. La création, modification ou suppression se fait via une application Java permettant de remplir les champs. Par contre, il est impossible de visualiser la donnée associée, la seule information d'accès renseignée est l'emplacement sur le disque réseau. Enfin, ce catalogue n'a pas vocation à filtrer les utilisateurs ayant des droits d'utilisation sur la donnée. C'est donc une solution incomplète par rapport aux attendus, néanmoins, l'existence de fiches de métadonnées répertoriées permettra de renseigner certaines données inconnues par ailleurs.

PEIGEO

La Plateforme d'Échange de l'Information Géographique à la Réunion (PEIGEO) propose un ensemble de géo services à destination du grand public et des professionnels de l'information géographique. On y trouve en consultation et dans certains cas en téléchargement les données du conseil général de la Réunion, du parc national... Les données sont cataloguées selon la norme INSPIRE, et interrogeables sur tous les critères évoqués dans le cahier des charges. Elle fournit tous les services définis plus haut pour la consultation des données

GeoMayotte

Géo Mayotte est la plate forme de catalogage et d'échange de données géographiques, de mutualisation et de valorisation de l'information cartographique à destination des acteurs géomatiques et du grand public. Cette plateforme entre dans le cadre d'INSPIRE et est financée par des acteurs publics (dreal et daaf entre autres) et réalisé par des prestataires privés.

Si l'on compare le catalogue interne du Cirad aux deux solutions Geomayotte et PEIGEO, le projet attendu se rapproche de ces derniers: une interface de visualisation et la possibilité de visualiser/télécharger depuis la plateforme. On peut aussi les désigner sous les termes de "portail": en effet, les organismes derrière ces plateformes ne sont pas producteurs de données, elles se contentent d'ouvrir l'accès à des données à un emplacement géographique. Considérant l'inventaire des données du Cirad, qui détient surtout des données qu'il n'a pas produit, c'est finalement assez proche de ce qu'on attend du projet. Ces deux plateformes sont ce qu'on appelle des infrastructures de Données Spatiales (IDS).

L'INFRASTRUCTURE DE DONNÉES SPATIALE

A ce stade, nous aurons besoin des éléments suivants

- un site web
- un serveur de données géographiques, qui va permettre de visualiser la donnée sur le site
- un catalogue de données, qui va nous permettre d'inventorier la liste des données et de réaliser des requêtes dessus.

Des serveurs cartographiques et des catalogues de données libres et gratuits sont disponibles. Donc il nous faut trouver un moyen de les faire communiquer. C'est une partie du travail réalisé par une IDS que d'assembler ces éléments entre eux.

Conceptuellement, l'IDS est d'abord une organisation qui repose sur des accords de partage, une coordination entre ses membres et des systèmes informatiques qui intègrent un ensemble de services: catalogue, service, logiciels, données, applications, page Web ⁶ L'objectif est l'interopérabilité entre les systèmes déjà existants et celui à mettre en place: les systèmes informatiques doivent pouvoir communiquer entre eux. Et c'est en fait l'objectif de la directive INSPIRE: établir une infrastructure de données spatiales à l'échelle européenne, ou l'IDS de chaque institution ferait remonter ses données à un catalogue d'ensemble. C'est donc tout à la fois une organisation humaine qu'un ensemble de système informatique lié entre eux.

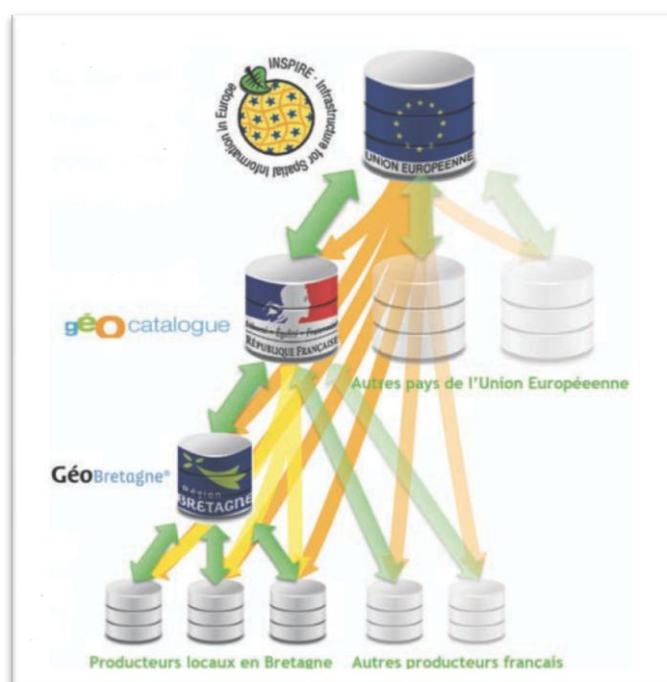


Figure 2 structure d'un IDS. source Geobretagne

Chaque IDS fait remonter une partie de ses informations à l'IDS de plus haut niveau. Ainsi le catalogue européen est censé tout centraliser. Dans l'organisation territoriale, une ville fait remonter à sa région, qui fait remonter à l'état, qui fait remonter à l'Union européenne.

Cette centralisation des systèmes de données géographiques est rendue technologiquement possible grâce à l'Open Geospatial Consortium (OGC): cette fondation à but non-lucrative a pour mission de favoriser l'échange de données géographique entre tous les acteurs de cet environnement. Pour ce

⁶ source wikipedia.fr

faire, l'OGC a développé des technologies informatiques désignées sous le nom de services: les OWS (Ogc Web Service), qui permettent de récupérer des flux d'informations géographiques à distance (sur un autre serveur web). Ces web services peuvent ensuite être consommés par un client web ou tout autre application les prenant en charge.

Avons-nous besoin d'une IDS pour ce projet ? Conceptuellement, le projet s'en rapproche: l'application sera consommatrice de services de visualisation, et le point d'entrée sera le catalogue, comme c'est le cas dans une IDS. Cependant, contrairement au réseau d'IDS présenté ci-dessus, il n'est pas envisagé pour l'instant de faire remonter des données aux systèmes de plus haut niveau: c'est un système qui se nourrit des données des autres acteurs, mais qui n'aura aucune donnée à communiquer aux autres IDS. Enfin, l'autre différence est qu'il n'est pas pour l'instant prévu d'aller récupérer des données sur des serveurs distants: Seules seront mises à disposition les ressources que possède le Cirad à la Réunion sur ses propres moyens de stockage.

INSTALLATION SUR SERVEUR (ALLOCATION RESSOURCES)

Créer une infrastructure de données spatiale opérationnelle à partir des briques logicielles "catalogue, serveur cartographique, site web" n'est pas possible en 6 mois de stage. Ce genre de système informatique existe déjà sous des formes de "cadre d'application", ou framework, qu'il faut déployer sur serveur et adapter aux ressources, aux données et aux besoins spécifiques de l'institution qui souhaite s'en doter. A ce stade, nous avons donc parcouru les Frameworks existants pour savoir lequel correspondait le mieux à ce projet d'Atlas. De cette veille technologique, sont remontées deux solutions, que nous allons présenter succinctement.

- EasySDI

Ce framework est développé par la société Depth, basée en Suisse et possédant une antenne à Toulouse. Son originalité est de reposer sur le Gestionnaire de contenu Joomla, pour l'administration des pages Web du site. Il est de conception modulaire, c'est à dire qu'on peut en théorie utiliser uniquement les parties du framework dont nous avons besoin (en l'occurrence, seulement l'aspect catalogue et le visualisateur).

- Georchestra

Cette solution logicielle a été développée par la Région Bretagne pour monter sa propre IDS. D'un seul tenant, sa communauté d'utilisateur est plus restreinte et l'ensemble des services proposés dépasse largement le cadre du cahier des charges. Le fait de devoir installer un système qui ne serait utilisé que très partiellement nous a dissuadé de l'essayer.

Nous avons donc entamé le déploiement d'EasySDI. L'allocation des ressources matérielles informatiques a été demandée à la direction des services informatiques, sur la base de la documentation d'EasySDI. Nous nous sommes concentrés sur trois aspects:

- La capacité de calcul de la machine (déterminé par le nombre de processeurs et leur fréquence de calcul) et la mémoire disponibles doivent permettre de faire fonctionner l'application correctement. Ce niveau minimal est déterminé à la lecture de la documentation d'EasySDI, sachant que l'élément qui requiert le plus de ressource est le serveur de données géographiques. dans le cas d'EasySDI, il s'agit de Geoserver. Nous avons donc demandé 4 processeurs cadencés à 2,6 Ghz et 6 Go de mémoire. Ces valeurs varient en fonction du nombre de connexions entrantes, et dans le cas de l'atlas, cette valeur reste évaluée à une quinzaine de connexions par jours, et

rarement plus de 5 connexions simultanées. Les capacités en mémoire pourront de toute façon être ajusté en fonction de l'évolution du besoin, après accord de la DSI

- L'espace de stockage sur un serveur web est aussi couteux pour la DSI. En chiffrant les données sur bebour à 42 Go, nous pensions que 100Go seraient largement suffisants, même si cela n'a pas été le cas: en effet nous avons rapidement dépassé ce plafond et demandé une extension d'espace disque
- Enfin, il faut un système d'exploitation de ce matériel. tout l'environnement d'EasySDI étant compatible GNU/Linux, et de plus, il nous faut rester dans le cadre de solution libre, nous avons donc choisi Ubuntu server 14.04

ECHEC AVEC EASYSDI , PASSAGE À GEONODE

Malheureusement, cette tentative de déploiement d'EasySDI s'est soldée par un échec. Assisté de l'ingénieur informatique Mickael Mezino pour l'installation, nous ne sommes pas parvenus, après avoir déployé tous les composants, à faire fonctionner le module de catalogage. Trois semaines ont été consacré à tenter ce déploiement.

La raison principale tient au fait que les versions de l'application déposée sur le site Internet officiel ne sont pas "fonctionnelles". Elles nécessitent une importante phase d'ajustement avant de fonctionner. Soit l'on a le temps nécessaire pour réaliser cette tâche préalable, soit il faut faire appel à la société Depth qui facturera une prestation pour l'assistance à l'installation. Cette assistance n'est pas comprise dans le cadre du stage, nous avons donc abandonné EasySDI. De cette impasse, il faut retenir que les développeurs de logiciels libre n'ont pas d'obligation de fournir un programme qui puisse être déployé en condition de production; parfois, les logiciels sont mis en licence libre avant tout pour bénéficier de la contribution d'autres développeurs, ou même, de façon plus implicite, dans une démarche commerciale. Après cette tentative de 3 semaines en avril, il nous fallait repartir sur une autre solution logicielle, sans savoir laquelle

Un autre membre du Cirad avait eu à déployer une Infrastructure de données spatiale. Vincent BONNAL, ingénieur en télédétection et membre du Cirad en Guadeloupe, avait rencontré les mêmes difficultés que notre équipe avec EasySDI. Nous donc avons pris contact par le biais d'un message qu'il avait posté sur un forum et avons appris qu'il s'était tourné vers la solution Geonode. C'est le logiciel qui a finalement été retenu. A partir de là, le développement de l'application a réellement pu commencer.

IV. DÉVELOPPEMENT

a. PRÉSENTATION DE GEONODE

L'application Geonode est une IDS dont le développement a commencé en 2010, par le biais de la société BoundlessGeo (anciennement OpenGeo), qui est la principale contributrice sur les projets Geoserver et PostGIS, logiciels utilisés dans Geonode. Les commanditaires initiaux sont le Programme Alimentaire Mondial (World Food Program, WFP). La présentation du concept d'IDS par les développeurs de Geonode diffère légèrement de celle avancée dans ce rapport :

"Les infrastructures de données spatiales fournissent une base à la découverte de données spatiales, d'évaluation de ces données, et d'application cartographiques pour les utilisateurs et les fournisseurs de données, à tous les niveaux des échelons gouvernementaux, du secteur commercial, des organismes à but non lucratifs, académiques et aux citoyens en général"

Cette définition est différente d'INSPIRE: Il s'agit ici d'offrir des outils d'exploration, mais aussi de contribution à la création de données géographiques. L'aspect communautaire est plus marqué: L'IDS vise ici une communauté d'utilisateurs-contributeurs qui travaillent pour le même organisme et qui vont utiliser la plateforme à des fins de partage. C'est l'outil idéal pour la création de données en **Crowdsourcing**. En un sens, cela correspond à la demande initiale: l'atlas du Cirad cible avant tout la communauté des agents du Cirad à la Réunion, même s'il n'est pas prévu de créer de la donnée directement via la plateforme Web.

Dès lors que ce projet a été choisi, nous sommes de nouveau passés à la phase d'installation. Il a été possible de communiquer avec la communauté de développeur du projet via la mailing list des contributeurs au développement de Geonode. Ces échanges par e-mail ont été riches d'enseignement et ont permis de faire avancer le projet plus rapidement. Le but des développeurs est que les gestionnaires d'IDS Geonode, une fois les plateformes mises en production, rapportent les bugs et mieux encore, reversent leurs éventuelles améliorations à la version officielle du projet

A la lecture de la documentation, Il a vite été clair que le projet, bien que fonctionnel et paraissant performant ne seyait pas exactement à la commande: en effet, d'une plateforme communautaire, où les utilisateurs connectés ont tous des droits identiques, il nous fallait passer à un portail institutionnel. Cela se traduit par un ensemble de modifications (entre autres passer d'une organisation dans laquelle les utilisateurs sont potentiellement tous contributeurs à une autre, avec des droits d'utilisation plus nuancés selon leur statut). Une fois l'installation réalisée, il a donc s'agit de transformer le logiciel pour l'adapter au besoin de la commande initiale. Ce sont de ces changements dont il va être question dans cette partie, ainsi que de la complétion du catalogue. Nous commencerons d'abord par présenter les éléments logiciels qui composent Geonode: le catalogue, le serveur de données géographique et l'interface web incluse.

FONCTIONNALITÉS

Au sein d'un catalogue, Geonode gère un ensemble de ressources: ce peut être des données de type jeux de données « vectorielles » ou raster (appelé couche dans le vocabulaire Geonode), et documents au format pdf ou word. A cela s'ajoute les « cartes », qui sont des compositions de plusieurs couches de données superposées, comme on pourrait le faire dans un logiciel SIG. Il est possible d'éditer les données vectorielles comme dans un logiciel SIG. Enfin on notera la possibilité d'éditer la sémiologie

des couches de données vectorielles. Il est même proposé d'exporter le résultat au format image ou PDF une fois la composition terminée. Ces fonctions supplémentaires, non incluses dans le cahier des charges, ont néanmoins été conservées pour quelques usages précis que nous verrons lorsque nous décrirons la complétion du catalogue.

LES BRIQUES LOGICIELLES FONDAMENTALES

L'IDS Geonode est complexe dans la mesure où ses composants logiciels sont nombreux. Geonode est basé sur les standards de l'OGC énoncés plus haut, les services OWS, qui permettent de servir des données géographiques sous divers formats dans un navigateur Web.

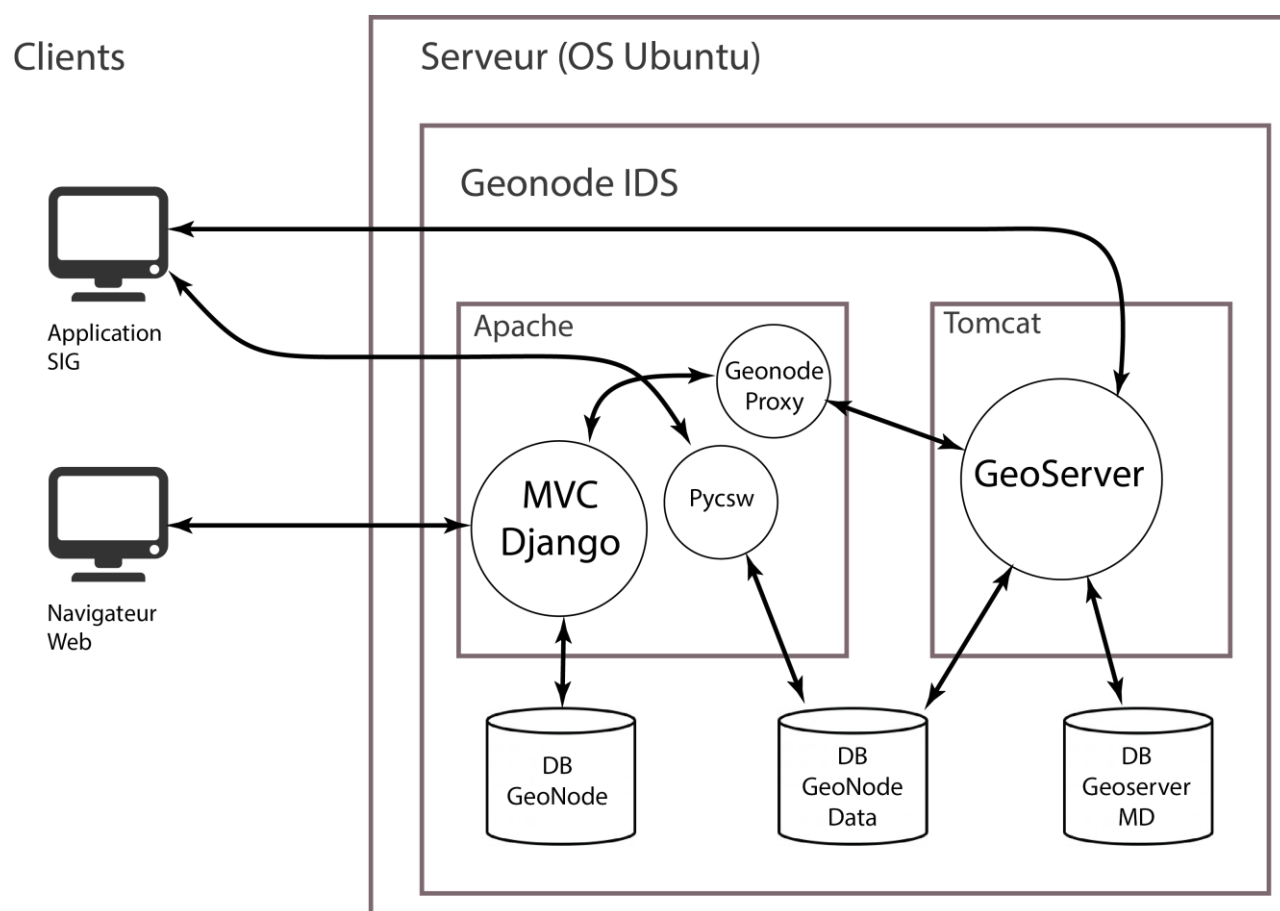


Figure 3 Schéma de la communication entre les briques logicielles

Revenons sur les 3 éléments principaux de ce schéma

DJANGO

Cadre de l'application Web de l'IDS. Codé en Python, il repose sur l'architecture dite "Model View Controller" (MVC): dans les échanges client-serveur d'une application, chaque requête du client est intercepté par le code du contrôleur, qui agit en fonction de cette requête pour construire la réponse, en interrogeant la base de données selon le modèle de base de données, et en construisant en guise de réponse une "vue", des données au format html que le navigateur du client sera à même d'afficher.

PYCSW

Module Python qui gère le catalogue de l'application. Il assure la prise en charge du service CSW de l'OGC. Si le catalogue est interrogé de l'extérieur, par une application comme Qgis notamment, c'est ce module qui est sollicité. En revanche, lors d'un filtrage sur les données via le site web, c'est bien l'API de Django qui dans un premier temps prend en charge la demande de filtrage. Donc ce module de catalogage n'est pas utilisé par le client web qui visite le site internet basé sur Django. Si c'est bien la même base de données qui est requêtée, les intermédiaires logiciels ne sont pas les mêmes.

GEOSERVER

Programme qui fonctionne dans la machine virtuelle Java via le conteneur de Servlet Tomcat. C'est lui qui est chargé de répondre à toutes les autres demandes de services OGC en provenance du client. C'est un serveur de données géographique intégrée comme un composant de l'IDS. Le proxy de Geonode en bride l'accès: Un client web pourra ne pourra y accéder que s'il est identifié via ce proxy, pour des raisons de sécurité. Ainsi il est impossible d'interroger directement Geoserver.

LES BASE DE DONNÉES DU SYSTÈME

il y a 3 bases de données Postgres différentes:

- Une base gérant les métadonnées, les autorisations, les utilisateurs, les permissions sur les ressources, et toutes les données nécessaires au fonctionnement de Django. cette base de données est appelée "Geonode"
- Une autre base de données avec en plus la cartouche postGIS pour les requêtes spatiales. Celle ci contient toutes les données vectorielles de l'application. A chaque table correspond une couche. C'est cette base qui en interaction avec Geoserver pour restituer les ressources.
- Enfin Geoserver gère ses propres métadonnées sur les ressources. Certaines sont en doublon avec les métadonnées présentes dans la base gérée par Geonode. C'est pourquoi il faut éviter les modifications directes avec un client de base de données, et passer par les interfaces d'administration de Geonode pour faire des modifications.

Si l'on se reporte au schéma d'organisation simplifié de la base de données Geonode ([Annexe B](#)), l'on peut voir les propriétés qui sont communes à tous tps de ressources dans la table resource_base. Les permissions sont gérées en associant une ressource, un utilisateur ou un groupe, et un type d'autorisation au travers d'une relation de type association.

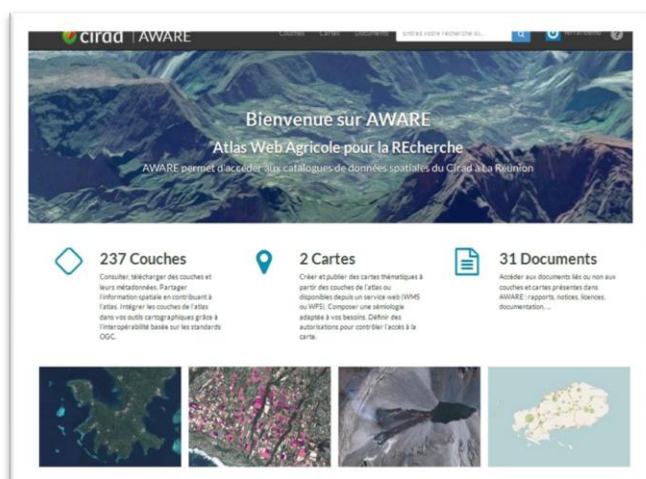
REPRODUCTION DES SCÉNARIOS D'UTILISATION ÉVOQUÉS DANS LE CAHIER DES CHARGES

Scénario 0 connexion d'un utilisateur

Ce scénario n'avait pas été annoncé dans le cahier des charges, mais la distinction entre utilisateurs s'effectue par l'authentification par login-mot de passe et ouvre les droits d'accès appropriés à l'identité du profil connecté.

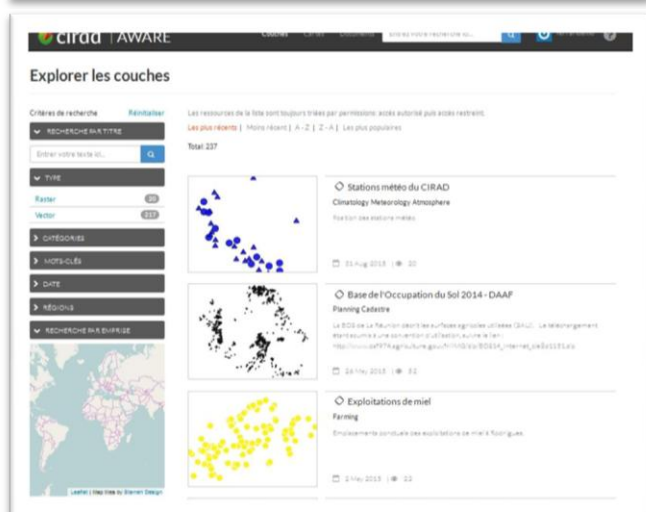
scénario 1 consultation

Le plan du site web Geonode est simple:



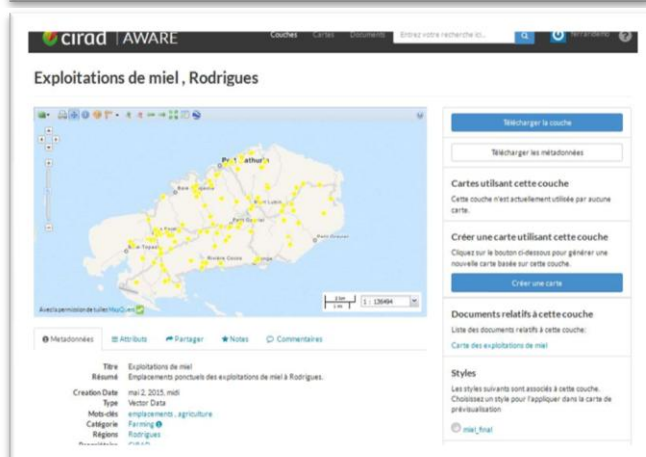
Ecran d'accueil

en point d'entrée, la page d'accueil. On sélectionne un type de ressource que l'on veut consulter, ou bien on fait une recherche textuelle sur la barre située dans le bandeau supérieur de l'écran auquel cas notre recherche portera sur tous les types de ressource



Liste des données•

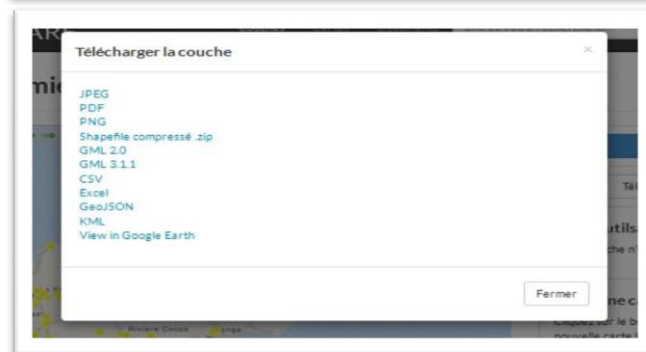
On accède à la liste des données correspondant à notre recherche, classé par défaut par date, du plus récent au plus ancien, bien que l'on puisse basculer à un tri par ordre alphabétique.



Détail d'une ressource

Après une sélection, on accède au détail d'une ressource, avec un visualisateur cartographique, et les métadonnées associées

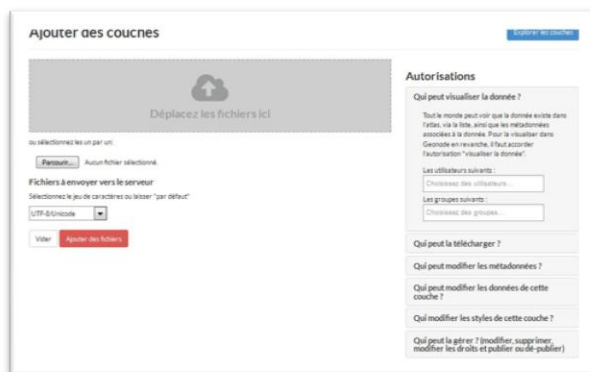
•



Liste des formats de téléchargement possible

Finalement, si l'on a l'autorisation, on peut choisir le format de téléchargement. Si l'on a pas l'autorisation, on peut adresser une demande au point de contact.

scénario 2 ajout de données supplémentaires



Ecran d'ajout de données

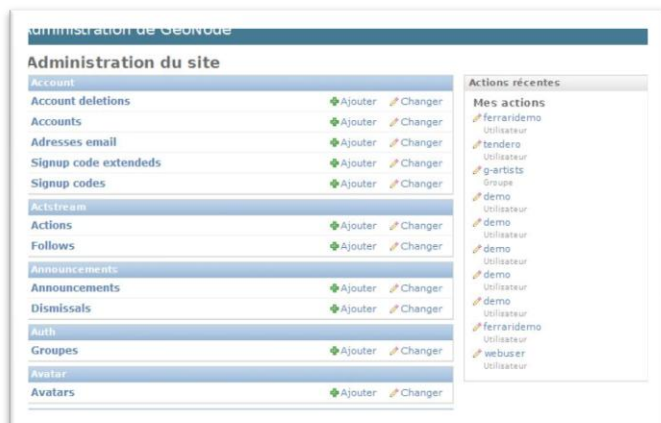
Au moment de l'ajout d'une donnée, on peut spécifier les autorisations accordées aux utilisateurs, sur lesquels



Modification des métadonnées

Après l'upload des données on peut accéder à l'écran de modification des métadonnées.

scénario 3 administration



Menu de gestion

Les gestionnaires du site peuvent accéder au back office (menu fourni avec Django). On peut y créer et supprimer des utilisateurs ou des groupes.

LES PERMISSIONS DANS GEONODE

Les permissions sont au cœur du projet car les problèmes de sécurisation des données sont gérées par leur biais. Geonode associe au couple utilisateur/ressource une série de permissions. Précisons qu'un utilisateur peut être inclus dans un groupe d'utilisateurs, auquel cas il hérite des permissions du groupe. Les différentes permissions d'un utilisateur/groupe sur une ressource sont:

- **voir la donnée:** l'utilisateur peut la voir dans la liste des ressources et peut la visualiser dans l'interface
- **télécharger la donnée:** l'utilisateur peut télécharger la donnée sur son poste.
- **modifier la métadonnée**
- **modifier la donnée :** Geonode bénéficiant d'un éditeur de données, on peut autoriser l'utilisateur à éditer la couche directement
- **modifier le style:** modifier la sémiologie de la donnée
- **gérer la ressource:** éditer les permissions, supprimer ou remplacer la couche

A l'usage, on a quasiment systématiquement regroupé les permissions « voir » et « télécharger », ainsi que « modifier la métadonnée », « modifier le style » et « gérer la ressource » qui sont fortement liées dans notre cas. La permission « modifier la donnée » n'est jamais accordée : il n'est pas envisagé pour l'instant d'éditer sur la plateforme. Précisons qu'un utilisateur peut être inclus dans un groupe d'utilisateurs, et dans ce cas il héritera des permissions allouées à son groupe.

b. ADAPTATION DE L'APPLICATION AUX NÉCESSITÉS DU PROJET

Si la suite logicielle standard permet de réaliser les scénarios imaginés au lancement du projet, le comportement de l'application n'est parfois pas celui attendu. Il faut donc apporter les modifications nécessaires.

OUTILS DE DÉVELOPPEMENT ET D'ADMINISTRATION SYSTÈME UTILISÉS

Généralement, lors d'un développement logiciel, on distingue l'environnement de développement qui facilite l'intégration ou la modification de code, de l'environnement de production, qui correspond aux conditions de fonctionnement réels du logiciel. Parmi les différences en développement web, la principale réside dans le fait de modifier le programme d'abord sur un serveur accessible uniquement aux programmeurs, pour ensuite l'intégrer sur le serveur accessible à tous. En effet, lorsqu'un programme est mis à disposition des utilisateurs (en production), on évite d'y intégrer des nouvelles modifications directement "sur site" pour éviter les interruptions ou des défaillances de service. Dans le cas de l'atlas, nous n'avons qu'un seul environnement: comme l'application n'a été mise en service qu'une fois complètement fonctionnelle, on a pu travailler directement sur le serveur d'hébergement final.

Ainsi, pour les commandes d'administration du système d'exploitation du serveur, j'utilisais un client SSH nommé putty, qui permettait de relancer le serveur apache ou tomcat lorsque nécessaire, par des commandes UNIX. Pour le développement, l'éditeur de texte Notepad++ possède un plug-in FTP qui enregistre directement les modifications de fichiers sur le serveur distant, sans avoir à passer par une application FTP comme Filezilla. Enfin, pour rendre compte des changements à l'équipe Artists, nous utilisons l'application Mantis Bug Tracker, qui permet de rapporter des bugs puis de notifier leur correction.

Le périmètre d'intervention sur le code est quasiment exclusivement limité au framework Django: du code Python, les templates en html « étendu » par les template tags et un peu de Javascript: nous n'avons pas apporté de modifications à Geoserver. Les modifications listées ci-après sont relativement triviales: elles ne nécessitent pas d'algorithmes complexes ou un grand nombre de lignes de code. La difficulté majeure vient plutôt de leur intégration à un environnement logiciel

inconnu au début du stage. Il a finalement été plus long d'appréhender le code existant que de développer.

On présentera dans les paragraphes suivants les modifications principales apportées, sous la forme suivante: l'état du logiciel avant intervention, la modification souhaitée et la méthode pour y parvenir, quand cela a été possible

INVENTAIRE DES DONNÉES PAR LA GESTION DES PERMISSIONS

Etat du logiciel avant intervention

Geonode calque son comportement sur celui de Geoserver pour l'affichage des ressources. Si un utilisateur n'a pas les droits d'accès à une ressource alors il ne pourra pas connaître son existence.

Modification souhaitée

L'utilisateur peut connaître l'existence de toutes les données, via l'écran de liste. S'il n'a pas les permissions nécessaires, on lui indique la personne à contacter pour demander le téléchargement. Il faut donc par défaut tout lister, et ensuite bloquer ou non l'utilisateur. Dans tous les cas, on veut que l'utilisateur puisse accéder à une fiche de métadonnées lisible. Sur la page d'accueil, on veut connaître le nombre de ressources totale, autorisées ou non pour l'utilisateur. Sur la page de listes, on affiche là aussi toutes les ressources disponibles, avec un verrou sur les ressources auxquelles l'utilisateur n'a pas accès. Et enfin, si l'utilisateur tente quand même de visualiser, on lui affiche un message lui indiquant la personne à contacter pour obtenir la permission.

Méthode utilisée

L'accès à la liste de toutes les ressources s'est fait en modifiant les instructions qui retournent la liste des ressources disponibles: plutôt que d'écarter les ressources sur lesquels l'utilisateur n'a pas droit, on renvoie toute la liste. Ensuite, si dans le code Python il était indiqué de renvoyer une erreur html 403 (accès refusé/interdit) à l'utilisateur essayant d'accéder à une ressource qu'il ne peut pas visualiser, l'on retourne à la place la même page qu'un utilisateur autorisé, excepté le cadre de visualisation et toutes les options de téléchargement ou modification.

GESTION D'UN GROUPE DE CONTRIBUTEURS

Etat du logiciel avant intervention

Chaque utilisateur connecté a l'autorisation d'ajouter des données dans Geonode.

Modification souhaitée

Le cahier des charges distingue le contributeur du visiteur. Les contributeurs sont certains membres des équipes de recherches du Cirad, et certains partenaires institutionnels. Une fois connecté, ces utilisateurs devront accéder aux écrans d'upload de données.

Méthode utilisée

On interdit l'accès à l'upload des données par défaut. Si l'utilisateur appartient au groupe des contributeurs, on l'autorise.

TRI DES DONNÉES SELON LES PERMISSIONS

Etat du logiciel avant intervention

Une fois effectué l'affichage de toutes les ressources disponibles, les données ne peuvent être triées que sur les critères de recherche: ordre alphabétique ou date. Les ressources sur lesquels l'utilisateur a des droits restreints et celles qu'il peut télécharger et visualiser ne sont pas ordonnées.

Modification souhaitée

On souhaite faire remonter en tête de liste les ressources sur lesquels l'utilisateur a des droits étendus de visualisation et téléchargement. Ainsi il verra d'abord les ressources qu'il peut utiliser, puis les autres.

Méthode utilisée

Modification de la réponse retournée par le serveur lorsqu'il renvoie la liste des ressources. Cette réponse, au format de listes Json est affichée dans le même ordre qu'elle est envoyée au client. Au moment de la construction de cette liste, on sépare les ressources en deux listes, une qui contient les ressources avec les permissions étendues et une autres avec les restrictions complètes. On met bout à bout les deux listes et on renvoie ce résultat.

EXPORT MÉTADONNÉES FORMAT HTML

Etat du logiciel avant intervention

Geonode propose, par l'intermédiaire de son module catalogue pycsw, de récupérer la fiche de métadonnées au format XML. Cela est un cas d'utilisation pertinent pour un utilisateur qui veut intégrer une fiche à son propre catalogue, si l'on ne veut pas tout *moissonner*. Pour un utilisateur lambda qui veut travailler la donnée sur son poste par contre, cela n'est pas pratique, car le format XML n'est pas lisible pour un humain.

Modification souhaitée

Proposer:

- un format texte importable dans un tableur comme Microsoft Excel, pour afficher chaque attribut de métadonnées dans un tableau.
- un format html pour visualiser et enregistrer directement depuis le navigateur web

Méthode utilisée

Reprendre la procédure d'export de la métadonnée au format XML, mais renvoyer une réponse formatée selon la demande utilisateur: HTML ou texte.

AUTHENTIFICATION PAR LE SERVEUR DE L'ANNUAIRE LDAP DU CIRAD

Etat du logiciel avant intervention

Django gère toutes les authentifications dans sa base d'utilisateur. Il faut ainsi créer un utilisateur manuellement et l'associer aux groupes auxquels il appartient.

Modification souhaitée

Le Cirad dispose d'une base de données de ses membres et des personnes ayant besoin d'un ou plusieurs services particuliers. Chaque utilisateur possède des attributs renseignant entre autres son lieu de travail et son organisme de rattachement. Cette annuaire permet d'accéder aux services par les mêmes identifiants de connexion, via un protocole particulier nommé Lightweight Directory Access Protocol LDAP.

Méthode utilisée

La prise en charge d'un annuaire LDAP par Django est possible via le plugin django-auth-ldap⁷. Plutôt que d'obliger un utilisateur à se créer un compte spécifique à l'application, on va vérifier s'il existe

⁷ On peut trouver la documentation sur pythonhosted.org

dans l'annuaire LDAP du Cirad. Si c'est le cas, il peut s'authentifier et est associé aux groupes auxquels il est rattaché.

ADAPTATION DE LA PAGE D'ACCUEIL ET MISE EN FORME GÉNÉRALE

Etat du logiciel avant intervention

La page d'accueil de Geonode, générique, ne représente pas le Cirad à la Réunion. Nous avons celle de l'application par défaut.

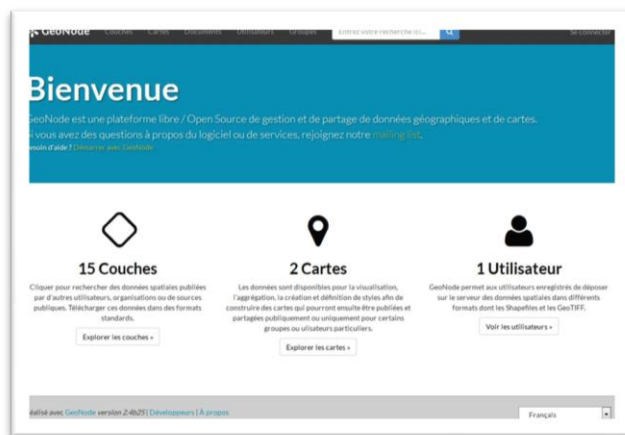


Figure 4 page de garde du site par défaut

Modification souhaitée

S'inspirer d'autres instances de Geonode, comme geonode.wfp.org ou masdap.mw.

Méthode utilisée

Le framework Bootstrap facilite l'adaptation de la mise en page à tous type de support: smartphone, tablette, ordinateurs de bureau. La nouvelle page de garde renvoie vers les sites partenaires. De la même manière de nombreuses modifications ont été ajoutés sur les gabarits, ou templates, de génération des pages HTML. La gestion de ces gabarits par Django permet d'ajouter de générer dynamiquement le contenu, à la différence d'une page HTML statique.

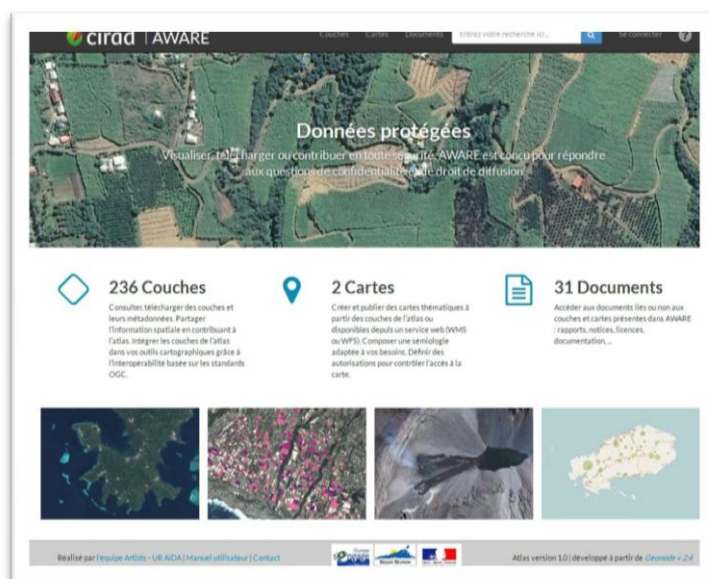


Figure 5 page de garde modifiée

MODIFICATIONS NON-TERMINÉES

Deux modifications supplémentaires étaient souhaitées en plus du cahier des charges, mais n'ont pu être menées à terme, par manque de temps.

Exploitation de Geoserver via un autre site

Les données gérées par l'instance de Geoserver auraient besoin d'être employées dans d'autres applications du Cirad: le volet cartographique de l'application margouilla notamment. S'il est bien possible d'importer des données par flux WMS/WFS depuis un logiciel SIG, en interrogeant directement Geoserver, il est impossible d'ouvrir des ressources dans l'atlas depuis une autre page web sans authentification préalable, car le proxy Geonode en interdit l'accès direct, pour la protection CSRF⁸. Geonode impose une authentification via Django si l'on tente d'accéder à Geoserver via un navigateur Web.

Import d'une sémiologie définie sur logiciel SIG

Après avoir définie une sémiologie d'une couche vectorielle via Qgis notamment, il eut été souhaité de l'intégrer à Geonode pour qu'elle soit restituée de la même manière. Seulement Qgis exporte des fichiers de style SLD de version 1.1, quand l'instance de Geoserver déployée avec Geonode n'accepte qu'en version 1.0. Il aurait donc fallu soit faire une mise à jour de version du geoserver embarqué, mais comme l'instance de Geoserver a été fortement adapté pour geonode, cela est déconseillé. J'aurais alors pu écrire un parser qui côté serveur, transforme la version 1.1 en 1.0, mais le temps a manqué. En l'état, la sémiologie doit être modifiée directement dans l'application. Tout import de fichier SLD en version 1.1 est perdu.

c. COMPLÉTION DU CATALOGUE

Une fois ces modifications terminées, il a fallu compléter le catalogue en ajoutant les ressources sur le serveur, puis en complétant les métadonnées associées. Cela nécessite de préparer la donnée avant son intégration. Cette étape est longue et répétitive mais a pu être simplifiée grâce aux fonctionnalités de Geonode permettant l'import multiple, et le fait d'avoir automatisé la complétion des champs de métadonnées quand ceux-ci étaient identiques au sein d'un jeu de données. On note aussi la possibilité d'importer un fichier XML contenant les métadonnées dans une des normes prises en charge par pycsw.

REPROJECTION PRÉALABLE

Geoserver accepte les données non-projetées et les données projetées, du moment qu'elles se basent sur le géoïde WGS84 (ou sur les géoïdes qui en sont issus, comme le RGR92). Les données de l'IGN à la Réunion sont par exemple fournies dans la projection Universal Transverse Mercator, fuseau 40, Sud (code EPSG :32740)., elles sont donc prises en charge directement. En revanche, les projections moins conventionnelles comme tananarive/laborde (code EPSG :29700) sur la zone de Madagascar, ne sont pas prises en compte, et dans ce cas, il faut déprojeter la donnée (sachant que la reprojection ne sera pas exacte, EPSG:29700 se basant sur un géoïde différent de WGS84).

DONNÉES VECTORIELLES

⁸ voir [documentation de Django](#)

Il faut obligatoirement les importer au format Shapefile, Geonode n'acceptant pas les fichiers Mapinfo. Cela n'est pas gênant car toutes les données du Cirad sont déjà dans ce format. Les fichiers de forme, ou shapefile, n'excédant pas 100 Mo, ils ont pu être intégrés par le processus d'upload intégré à Geonode. La donnée vectorielle est transformée en données PostGIS pour intégrer la liste de ressource de Geoserver.

DONNÉES RASTER

Pour les rasters, le processus est plus compliqué. En général, les fichiers sont trop volumineux pour être intégrés en upload par l'interface de Geonode. De plus les rasters sont souvent fragmentés en plusieurs fichiers qui composent une dalle. Assembler ces dalles en une seule alourdirait encore ces fichiers, le plafond indiqué par la documentation de geoserver étant 2Go. Les formats de ces fichiers sont geotiff, (pris en charge nativement par Geoserver) et ECW.

Pour l'ECW, s'est posé un autre problème: ce format n'est tout simplement pas supporté par geoserver, du moins par ses bibliothèques natives. De surcroît l'ECW est un format propriétaire, et le mettre en visualisation sur serveur nécessite de payer une licence d'utilisation. Nous avons donc fait le choix de convertir les fichiers ECW en GeoTiff, afin d'utiliser uniquement des technologies libres. La difficulté étant de conserver un volume de fichiers équivalent à l'ECW. Dans un deuxième temps, pour que Geoserver serve ces fichiers volumineux rapidement, nous avons construits des « pyramides » c'est-à-dire des mosaïques d'images à différents niveaux de zoom. Ainsi, plutôt que de renvoyer la dalle complète, Geoserver renvoie uniquement la section sur laquelle l'utilisateur est situé.

Ces deux traitements ont été réalisés via GDAL (Geospatial Data Abstraction Library), la bibliothèque libre permettant de lire et de traiter un très grand nombre de formats d'images géographiques. GDAL est utilisé directement en ligne de commande, et les opérations peuvent être chaînées.

Évidemment, la création de ces pyramides rend la donnée beaucoup plus volumineuse : Les 100 Go initialement alloué pour le système complet ont été dépassés. Il a donc fallu solliciter à nouveau la Direction des Services Informatiques pour obtenir 100 Go supplémentaires, et consacrer ce nouveau disque au stockage unique des données Raster.

COMPLÉTION DES MÉTADONNÉES

En parallèle de l'ajout d'une ressource, il faut compléter une fiche de métadonnées. Lorsqu'elle n'est pas présente sur le serveur, il faut la débusquer sur le site officiel des producteurs de données. C'est lors de la complétion de ces fiches que l'on prend conscience de l'intérêt de l'interopérabilité: Une fiche de métadonnées au format XML, et répondant à la norme ISO, pourrait être ajoutée durant la phase d'upload. Mais malheureusement, les fiches de métadonnées sont rarement complètes et encore moins normées. Dans le cas des BD TOPO de l'IGN, les fiches de métadonnées sont au niveau de l'ensemble de la Base de données. La granularité est trop faible, il faut donc aller récupérer les informations dans le descriptif au format PDF.

Il fallait donc saisir toute les fiches manuellement, une par une. Heureusement, la saisie a pu être à moitié automatisée en complétant pour chaque série de données, les métadonnées communément partagées (par exemple, pour les couches de la bdtopo : édition, date, thématique, propriétaire, entre autres). Pour ce faire, le gabarit (template Django) correspondant au formulaire de complétion d'une fiche de métadonnées a été agrémenté d'un script javascript qui complétait automatiquement

certaines champs. Cette semi automatisation mériterait d'ailleurs d'être généralisée à Geonode, afin d'obtenir des prototypes de fiches déjà partiellement complétées.

V. RÉSULTAT

a. BILAN QUANTITATIF

EN NOMBRE DE RESSOURCES

A la fin du stage, il y avait dans l'atlas

- 236 couches de données
 - 216 couches vectorielles
 - 20 rasters
- 30 documents au format PDF
 - 3 cartes
 - 25 descriptifs
 - 2 fichiers de licence
- 2 compositions de couches (ou « cartes Geonode »)

ZONES GÉOGRAPHIQUES

Les zones géographiques couvertes par ces données sont

- 193 couches sur L'île de la Réunion
- 33 couches sur Madagascar
- 6 couches sur le Kenya
- 3 couches sur l'île de Rodrigues

PRODUCTEUR DE DONNÉES

- Les ressources en provenance de l'IGN sont les plus nombreuses, avec 87 ressources
- Le Cirad compte 12 ressources

VOLUME DES DONNÉES

- on compte 92 Go de données raster pyramides incluses. Nous n'avons pu inclure que les ortho photographies, les France Raster et les SCAN de l'IGN, plus quelques scans de cartes sur Madagascar. L'imagerie satellitaire utilisée par l'équipe Artists n'a finalement pas été ajoutée, par manque de temps: sans inventaire et sans réflexion préalable sur la manière d'inclure cette donnée, beaucoup trop volumineuse par ailleurs, elle ne sera accessible que sur le serveur Artists.
- Les données vectorielles comptent quant à elles seulement 1,6 Go, alors que les fichiers shapefile remontés vers le serveur représentaient 4,2 Go. Cela est dû au fait que PostGIS

compresse les données shapefile en n'enregistrant seulement des topologies, même si elles sont restituées sous forme d'objet ensuite.

- Si le premier disque de 100 Go alloué au système a été insuffisant, c'est que l'espace nécessaire pour stocker le système lui-même est conséquent : tomcat7, qui est le conteneur de Geoserver, pèse à lui seul 41 Go. Les backups automatiques du code pèsent 15 Go supplémentaires. L'inventaire sera bientôt peuplé des contributions des autres équipes de recherche. Cependant, vu le faible volume occupé par les données shapefile convertis en PostGIS, on peut estimer que le disque principal ne devrait pas arriver à saturation rapidement.

b. RAPPORTÉ AU CAHIER DES CHARGES

Par rapport aux attendus, l'atlas remplit sa fonction d'inventaire, permet aux contributeurs de gérer leurs ressources, et à l'administrateur d'avoir la main sur l'ensemble de l'application. Le cahier des charges est donc rempli, au moins théoriquement : en effet l'on ne peut juger d'une application que lorsqu'elle a été adoptée par les utilisateurs auxquels elle est destinée : Y-a-t-il des manquements, des fonctionnalités qui font défaut? Un atelier de travail avec 2 futurs contributeurs (UMR PVBMT et UR Hortsys) a eu lieu, le 7 Mai, alors que l'application était en cours d'ajustement. En présence de chercheurs et futurs contributeurs, nous avons pu mesurer la différence entre l'utilisation par un habitué de l'application qui répète des centaines de fois les scénarios d'utilisation, et une personne moins expérimentée en SIG, qui ne l'a jamais utilisée: les défauts et les éventuels bugs apparaissent plus facilement.

Lorsque l'on est au cœur du développement d'un projet, il n'est pas évident d'avoir le recul nécessaire pour savoir ce qui sera pertinent ou au contraire, gênant pour l'utilisateur, en terme d'ergonomie notamment. Cette remarque s'ajoute au fait qu'il n'a pas été possible de former les membres du Cirad au produit final, comme initialement prévu. Ces séquences de formation auraient fait surgir des remarques et critiques pour ajuster l'application dans sa forme définitive. Agnès Tendero et Mickael Mézino ont cependant anticipé en faisant remonter les défauts ou bugs, qui ont pu être corrigés dans le temps imparti.

c. RAPPORTÉ AUX AMÉLIORATIONS SOUHAITÉES

L'ensemble des données n'a pu être intégré. Il manque notamment certaines données d'archive, comme la BDTOP0 2003 de l'IGN. De plus les données du serveur Artists n'ont pas été intégrées : même si leur volume excédait largement les capacités de stockage des serveurs de l'application, il aurait été possible de les échantillonner. Cette carence en image satellitaire a cependant une portée limitée: contrairement aux données SIG ajoutées dans l'atlas, qu'un utilisateur peu expérimenté en SIG peut employer, les compétences en télédétection nécessaire pour traiter les images radar ou pléiades, sont exclusives à l'équipe Artists, et le besoin extérieur est faible, voire nul.

Aussi, si l'on imagine un système avec de multiples contributeurs, il est possible que certains fassent preuve de négligence en remplissant la fiche de métadonnées ou remplissent des métadonnées incorrectes : un circuit de mise en quarantaine de la donnée, avant validation par les administrateurs de l'équipe Artists, qui la publie définitivement, aurait été préférable. A ce stade, tous les contributeurs publient de la même manière.

LA PROPOSITION DE MOISSONNAGE DES CATALOGUES EXTERNES

Nous avons constaté durant la phase d'inventaire, qu'une large part des données de l'atlas étaient produites à l'extérieur du Cirad: IGN, BRGM, DEAL ont tous leur portails de données géographiques. Reprenons une proposition formulée au départ : n'aurait-il pas été plus pertinent d'aller moissonner les catalogues des bases de données des autres fournisseurs, plutôt que de les stocker toutes leur données sur les serveurs du Cirad ? Geonode prend en charge cette fonctionnalité, qui est à double tranchant : cela permet d'avoir la donnée la plus à jour possible, puisque directement issu du serveur du fournisseur. Par contre, cette solution a des limites : les serveurs de données géographiques des autres fournisseurs ne sont parfois pas complètement interoperables ou performants (nous avons essayé de moissonner celui du BRGM notamment, et les données restituées étaient incomplètes). Le corolaire de cette méthode, c'est qu'elle rend l'atlas dépendant du bon fonctionnement des serveurs externes, qui sont pour certains situés en métropole, ce qui rallonge les temps de réponse par rapport à un utilisateur situé à la Réunion, or, c'est sur l'île que se situent l'essentiel des utilisateurs. Nous avons donc uniquement servi les données déjà stockées sur les disques du Cirad.

Dans le même élan, l'atlas gagnerait à documenter plus amplement les fournisseurs de données géographiques, qui sont finalement peu renseignés puisque la fiche de métadonnées se contente d'un nom de producteur: on pourrait imaginer une page par fournisseur, qui référencerait les liens utiles (les portails de chaque fournisseur par exemple). On se rend compte qu'il est parfois difficile de savoir où se trouve la donnée OpenData, et qu'orienter l'utilisateur sur les sites portail leur ferait gagner du temps de recherche.

d. BILAN EN TERMES DE GESTION DE PROJET

Si l'on se rapporte au Gantt que j'ai réalisé en début de stage ([annexe D](#)) et proposé à l'équipe, les phases du projet sont assez similaires à celles effectivement réalisées. On anticipait au début le développement logiciel complet de la plateforme, ce qui était en fait impossible en 6 mois de stage : nous sommes donc partis d'une solution existante, Geonode, pour l'adapter. L'aspect communication et formation des utilisateurs sur le produit final n'a pas pu être achevé car la plateforme est devenue fonctionnelle uniquement en fin de projet, à une période où ses utilisateurs sont difficilement mobilisables (fin août).

La communication avec les membres de l'équipe Artists, que je devais entretenir tout au long du stage, pour rendre compte du travail accompli et des difficultés rencontrées, a parfois été négligée de mon côté: il m'est arrivé de présenter des échéances que je n'ai pas tenu, et ne les justifiant pas, cela a ralenti l'avancé du projet. Ayant acquis une certaine autonomie dans mon travail, j'en oubliais les impératifs de compte rendu de mes activités, et cela est d'autant plus dommageable que les problèmes étaient la plupart du temps résolus plus vite en équipe. J'en retiens l'importance d'un échange clair et continu, et que l'autonomie acquise dans le cadre d'un projet n'est pas synonyme d'indépendance vis à vis des autres parties prenantes du projet.

Deux tâches chronophages ont entravé la progression du développement de l'atlas: l'installation et la tentative de déploiement de la solution EasySDI d'abord: entre le 27 mars et le 15 avril, presque 3 semaines auront été consacrées à tenter de déployer cette suite, sans succès. Il a donc fallu repartir de zéro. Ensuite la mission à Rodrigues, dans le cadre du projet Qualireg, en plus de la semaine

passée sur place (initialement prévue dans le planning) a demandé un temps de préparation en amont et en aval, plus épars et donc difficile à estimer.

VI. CONCLUSION

Le projet, dans sa dimension technique, est arrivé à maturité: l'Infrastructure de Données Spatiale est fonctionnelle, peut desservir les données en visualisation et en téléchargement. Il permet de distinguer la communauté d'utilisateurs de celle des contributeurs. Reste désormais à inciter les contributeurs potentiels à le prendre en main. La phase de communication à travers le Cirad sera entamée après le stage et car l'outil n'aura d'intérêt que s'il est adopté par les agents en quête de données géographiques.

L'équipe Artists, qui a désormais à sa charge la plateforme, s'occupera de cette étape de prospection. Le nom **AWARE**, acronyme de "**A**tlas **W**eb **A**gricole pour la **R**Echerche", a été choisi par sondage auprès des membres de l'équipe, stagiaires inclus. Il a été privilégié au détriment de noms réunionnais pour ne pas limiter son appropriation aux utilisateurs situés sur l'île. Cela correspond aux objectifs initiaux: fournir une plateforme de mutualisation et de partage qui pourra bénéficier à l'ensemble des utilisateurs au travers de l'océan Indien

De mon point de vue, l'intérêt de ce stage est d'avoir pu mener le projet de la phase préliminaire de détermination du besoin jusqu'à la mise en production au moment de mon départ. Cela permet d'avoir un condensé des écueils dans lesquels j'ai pu tomber dans ma gestion du projet. Au-delà des compétences techniques nécessaires à la mise en œuvre, il faut de nombreuses qualités supplémentaires de prise de recul et de synthèse, dont on s'écarte lorsque l'on est confronté à des problèmes très précis. Avoir eu la liberté de choisir les solutions possibles lors de la phase de maîtrise d'ouvrage m'a fait prendre conscience de l'importance d'une veille technologique rigoureuse avant la prise de décision.

VII. ANNEXES

a. VOLUMES DE DONNÉES SUR LE SERVEUR BEBOUR: TABLEAUX

Classement selon l'emplacement

Base de données Margouilla	2,6 Go
Disque réseau Bebour	42,6 Go
Serveur Artists	> 2600,0 Go

On ne prend en compte que les données sur Bebour, seule source de données prise en compte lors du premier inventaire.

Classement par organisme de référence

Organisme de référence	Volume en Go
Cirad	0,7
Conseil Régional	0,5
Conservatoire du littoral	0,2
DAAF	0,1
DEAI	0,4
IGN	36,4
ONF 2004	0,02
PARC	0,18
Sources diverses	2,9

L'IGN compte pour 87 % des données en présence. IL est à noter que la part des données provenant du Cirad va augmenter à mesure des contributions des chercheurs

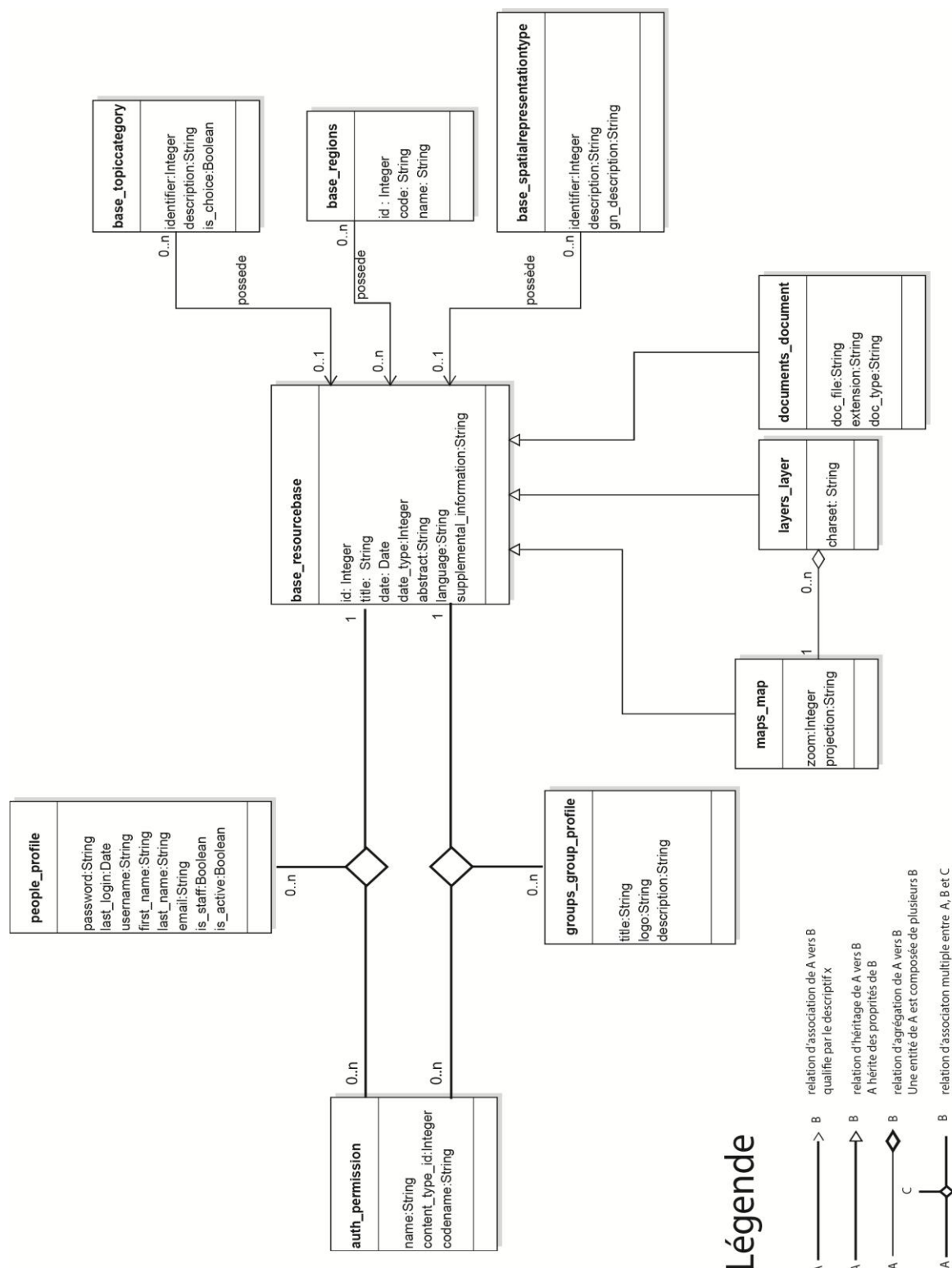
Classement par type de données

Type de fichiers	volume
Images pixels non géoréférencées (.jpg ou autre)	0,1
Pdf	0,3
tiff	4
Tab	0,3
shp	7
.ecw	36,4

Le format ecw est le plus volumineux, le bases de données ortho de l'IGN étant stockées dans ce format.

Il à noter que le volume de données finalement intégré à l'atlas est plus important car au moment ou l'inventaire a été réalisé, il manquait certaines bases de données présentes dans l'Atlas.

b. SCHÉMA SIMPLIFIÉ UML DE LA BASE DE DONNÉES "GEONODE"⁹



⁹ Pour réaliser ce schéma, nous il a fallu récupérer la strcture de la base de données via un dump sur Postgres, sachant qu'elle compte 74 tables. Après avoir tenté une opération de reverse engineering sur Win Design, n'ont été retenu dans ce schéma que les tables de gestion des ressources, utilisateurs et permissions, sans ajouter les tables des éventuels plug in.

c. UN CAS D'UTILISATION DE L'ATLAS À RODRIGUES

Dans la semaine du 31 mai au 6 juin, j'ai été envoyé sur l'île de Rodrigues réaliser une mission dans le cadre du projet Qualireg, qui vise la mise en place de démarches qualités officielles pour la valorisation d' agro-produits. Certains, en provenance, de Rodrigues ont été identifiés comme présentant une forte typicité et des attributs de qualité originaux : le miel, le piment et le citron-galet ete autres.

Le but de la mission était de compléter les données nécessaires à la complétion d'un des objectifs: construire une base de connaissance partagée sur les produits et les filières de Rodrigues les plus avancées. Pour remplir cet objectif, il est prévu de réaliser une plateforme de webmapping, pour l'identification des producteurs et opérateurs des filières illustrant leurs pratiques et leurs différents statuts au regard des informations géographiques. L'atlas jouerait alors ce rôle.

Il a donc fallu aller à Rodrigues pour présenter le projet d'Atlas aux représentants de la Commission de l'Agriculture de Rodrigues, partenaire du projet. La commission accepterait de partager ses données géographiques (une collection de données topographiques: réseau routier, zones administratives, ect).

Ces données, vectorisées par un agent de la commission à partir d'une image satellitaire Quickbird mal géoréférencées, n'ont pu être intégrées, du fait du décalage des coordonnées saisies avec les coordonnées réelles.

De plus, il a été prévu de réaliser une carte des positions de chaque exploitation agricole pour les produits des filières désignées. En plus des positions des exploitations, on récupère le volume de production de chaque exploitant. De cette manière, on peut connaître les zones majoritaires de production sur l'île et éventuellement améliorer les réseaux de distribution. Les données n'ayant pas été entièrement collectées pour la filière piment, j'ai réalisé plusieurs jours de relevés par GPS. Il s'agit de relevés ponctuels, la précision à 5 mètres du GPS ne permettant pas de définir le contour précis de parcelles parfois inférieures à 10m². Une fois l'intégralité des relevés d'une filière accomplie, il était possible de réaliser un assemblage de ces données en shapefile et d'en faire une carte « finale » avec une représentation par symbole proportionnel du volume de production.

L'intérêt de ces cartes est limité dans le temps car les exploitations changent fréquemment de place, en particulier pour la filière piment. C'est donc une donnée qui devra être fréquemment mise à jour. Dans le processus d'utilisation de l'atlas, un contributeur pour les données Rodriguaises peut être désigné à cette tâche. Un groupe d'utilisateurs peut avoir accès à ces données en visualisation & téléchargement, sans avoir accès aux autres données du Cirad.

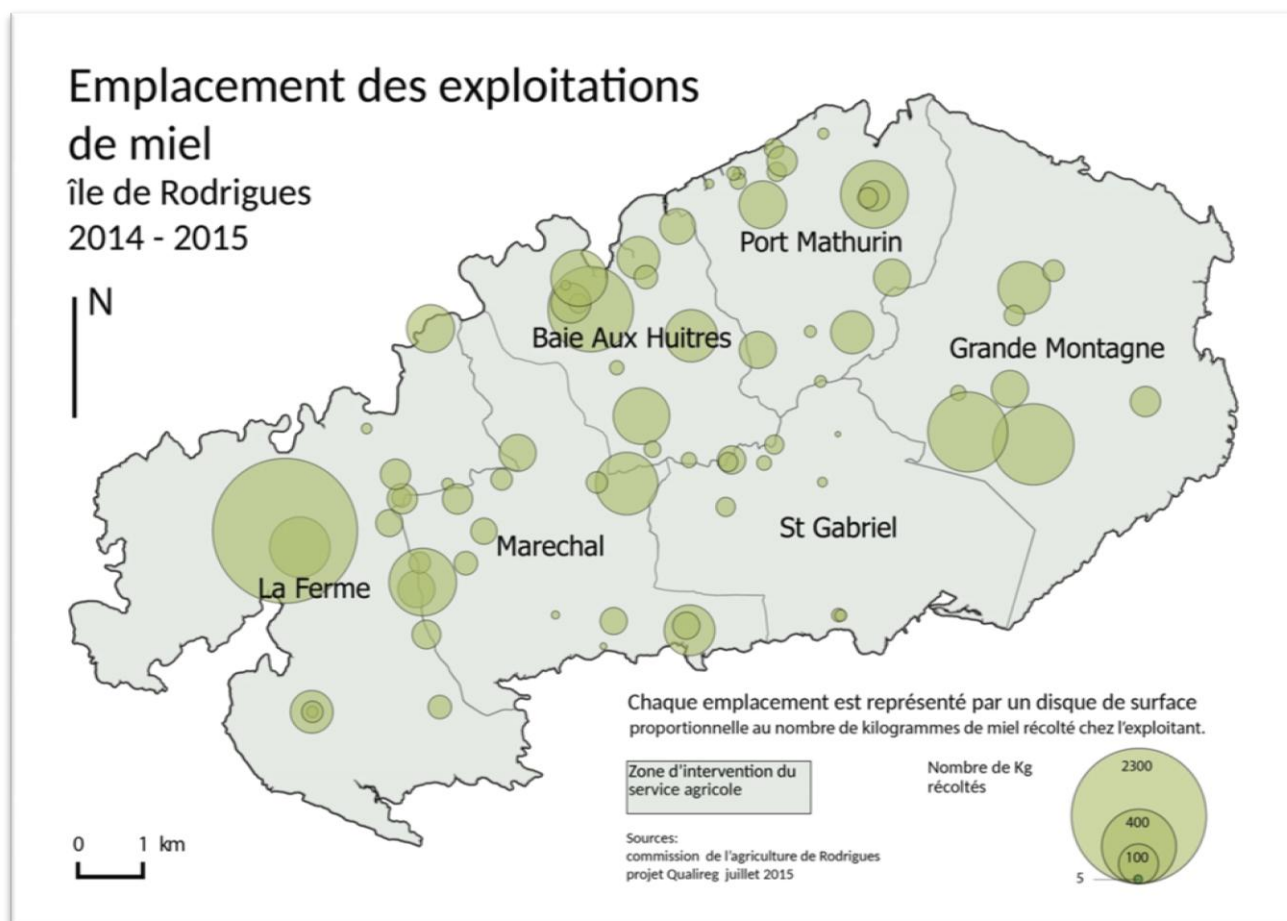


Figure 6 exemple d'une carte pour la filière miel à Rodrigues

Ces données ont été diffusées dans l'atlas de deux manières différentes : sous la forme de carte définitive au format Pdf, comme on le voit dans la figure ci-dessus. Mais aussi sous forme de couches de données, ou cette fois la sémiologie est identique pour tous les emplacements. La couche de données pointe sur la carte, et le visiteur qui verra la donnée pourra être redirigé sur la carte et inversement.

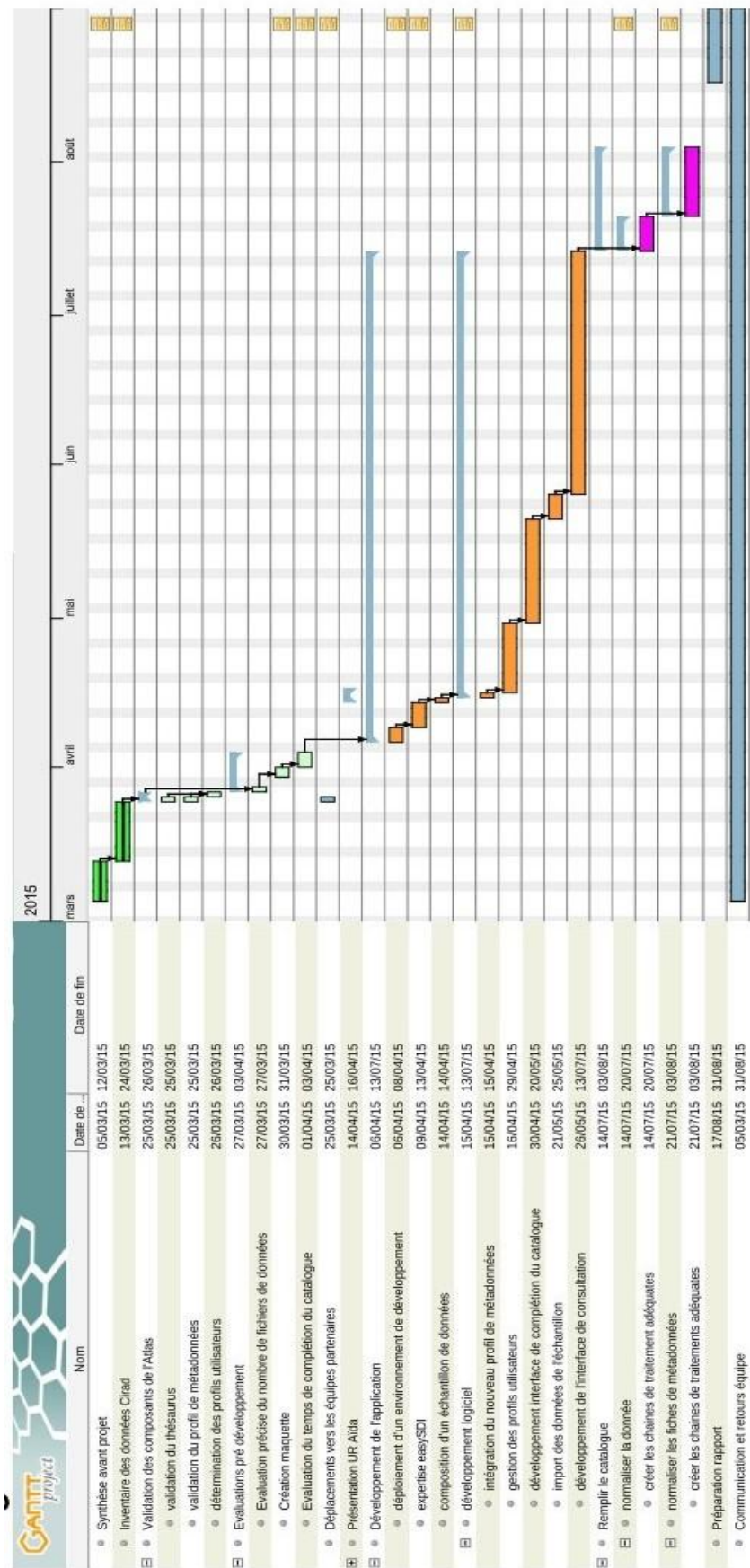
Nous touchons là une autre limite de Geoserver: cette sémiologie par symbole proportionnel n'est pas encore prise en charge par l'application. Nous ne pouvons encore diffuser une couche de données par symbole proportionnels directement, d'où cette double diffusion.

d. PLANNING DE GANTT, RÉALISÉ EN DÉBUT DE PROJET (15 MARS 2015)¹⁰

Tâches

Nom	Date de début	Date de fin
Synthèse avant projet <i>définir les objectifs et le cadre du projet présenter les données et services existants proposer des solutions techniques répondant à la commande</i>	05/03/15	12/03/15
Inventaire des données Cirad <i>Bilan de l' inventaire sur bebour et artist Composition de la fiche du futur catalogue Etat de l'art des applications de catalogage existantes et des framework de création d'IDS.</i>	13/03/15	24/03/15
Validation des composants de l'Atlas	25/03/15	26/03/15
validation du thésaurus	25/03/15	25/03/15
validation du profil de métadonnées	25/03/15	25/03/15
détermination des profils utilisateurs	26/03/15	26/03/15
Evaluations pré développement	27/03/15	03/04/15
Evaluation précise du nombre de fichiers de données	27/03/15	27/03/15
Création maquette <i>Expliciter les attendus de l'interface utilisateur à l'aide de "visuels" de ce à quoi elle pourrait ressembler avancer sur le contenu du site, en dehors du catalogue: quid de la partie vitrine et présentation du cirad à la Réunion ?</i>	30/03/15	31/03/15
Evaluation du temps de complétion du catalogue <i>Si la tâche "remplir le catalogue" suit le développement de l'application, il faudra préciser en détail quel est le travail à fournir et le temps nécessaire pour compléter le catalogue, avant de commencer le développement</i>	01/04/15	03/04/15
Déplacements vers les équipes partenaires <i>Plusieurs déplacements prévus auprès des différentes antennes à Saint Denis notamment. Les dates ne sont pas encore fixées, il s'agit d'une date par défaut.</i>	25/03/15	25/03/15
Présentation UR Aïda	14/04/15	16/04/15
préparation présentation	14/04/15	15/04/15
journée de présentation	16/04/15	16/04/15
Développement de l'application	06/04/15	13/07/15
déploiement d'un environnement de développement <i>prendre en compte l'environnement de production. estimer le volume de développement global</i>	06/04/15	08/04/15
expertise easySDI <i>S'assurer des possibilités d'easySDI en terme de visualisation et de complétion de catalogue, une fois l'environnement de développement posé.</i>	09/04/15	13/04/15
composition d'un échantillon de données	14/04/15	14/04/15
développement logiciel <i>regroupe toutes les tâches ayant trait au développement des cas d'utilisation proprement dits.</i>	15/04/15	13/07/15
intégration du nouveau profil de métadonnées	15/04/15	15/04/15
gestion des profils utilisateurs	16/04/15	29/04/15
développement interface de complétion du catalogue	30/04/15	20/05/15
import des données de l'échantillon	21/05/15	25/05/15
développement de l'interface de consultation	26/05/15	13/07/15
Remplir le catalogue	14/07/15	03/08/15
normaliser la donnée <i>déprojeter convertir en shape placer dans la nouvelle arborescence du serveur</i>	14/07/15	20/07/15
créer les chaines de traitement adéquates	14/07/15	20/07/15
normaliser les fiches de métadonnées <i>automatiser la récupération des métadonnées existantes remplir à la main quand c'est impossible</i>	21/07/15	03/08/15
créer les chaines de traitements adéquates	21/07/15	03/08/15
Préparation rapport	17/08/15	31/08/15
Communication et retours équipe	05/03/15	31/08/15

¹⁰ Réalisé avec le logiciel GANTT Project



e. ACRONYME ET GLOSSAIRE

ACRONYMES

Artists: Artists of Remote sensing Tools, Information systems, Simulation Techniques and Spatial analysis

Aïda: Agroécologie et Intensification Durable des Cultures Annuelles)

ECW: Enhanced Compressed Wavelet

EPIC : Etablissement à caractère industriel et commercial

EPSG: European Petroleum Survey Group

Json: Javascript Object Notation

UR : Unité de recherche

Persyst: Performance des systèmes de production et de transformation tropicaux

XML: eXtended Markup Language

SSH: Secure Shell - protocole de communication sécurisé

GLOSSAIRE ET DÉFINITIONS

Logiciel Libre

Logiciel dont l'utilisation, l'étude, la modification et la duplication en vue de sa diffusion sont permises, techniquement et légalement. Ceci afin de garantir certaines libertés induites, dont le contrôle du programme par l'utilisateur et la possibilité de partage entre individus. L'initiative en revient à Richard Stallman et la [Free Software Foundation](#)

OpenData (ou donnée ouverte)

Donnée numérique d'origine publique ou privée. Elle peut être notamment produite par une collectivité, un service public (éventuellement délégué) ou une entreprise. Elle est diffusée de manière structurée selon une méthode et une licence ouverte garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.

Crowdsourcing

Utilisation de la créativité, de l'intelligence et du savoir-faire d'un grand nombre de personnes (contributeurs), en sous-traitance, pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur. Des exemples significatifs sont le projet OpenstreetMap, qui est une carte géographique à échelle mondiale vectorisée par des volontaires via une plateforme web interactive. La cartographie en période de crise à Haïti en 2010 par exemple, est souvent réalisée de la sorte, par des bénévoles soucieux d'aider les équipes humanitaires sur place.

Représentation de données géographiques

Généralement, on sépare en deux catégories la donnée géographique: la représentation vectorielle associe à chaque entité du monde réel un objet: un point une ligne ou un polygone délimitant une surface, qui sont référencés dans l'espace. A l'inverse, il existe la représentation de type "grille"(généralement grille de pixels de taille uniformes, ou raster), où l'espace représenté est découpé en unités spatiales régulières, selon une trame. Il existe d'autres manières de représenter la donnée spatiale mais elles n'ont pas été répertoriées dans les données du Cirad.

Shapefile

Développé par la société ESRI, ce format est devenu un standard, et est désormais supporté par toutes les applications, libres ou propriétaires. C'est aussi le plus couramment utilisé dans les applications SIG. L'avantage est qu'il associe à chacun de ces objets des données attributaires, c'est à dire des informations qui décrivent l'objet lui-même.

Geotiff

Standard du domaine public permettant d'ajouter des informations de géoréférencement à une image TIFF (projection, système de coordonnées, datation, ...)

Moissonnage de métadonnées

Opération consistant à extraire automatiquement des métadonnées d'un catalogue pour l'ajouter à un autre.

Template (gabarit) Django

Un gabarit contient la partie statique du résultat HTML souhaité ainsi qu'une certaine syntaxe particulière définissant comment insérer le contenu dynamique.

Remerciements

Voici encore un inventaire, le dernier de ce rapport. je remercie:

Les contribuables français qui ont financé mes études. Merci à vous, j'espère vous rendre la pareille rapidement.

Richard Stallman et la Free Software Foundation, pour avoir inventé un système de partage des connaissances basés sur des principes que je cautionne.

L'équipe de développement de Geonode pour avoir respecté les principes évoqués ci dessus.

L'équipe Artists, dont Agnès Tendero, Mickael Mezino pour la qualité de leur encadrement, Pierre Todoroff et Louis Paulin.

Laurent Jégou, et l'équipe du Master SIGMA

Wikipédia, pour toutes les connaissances et quelques erreurs ça et là, rien de grave cependant.

Vincent BONNAL, qui a trouvé la solution. Lucien Weaspear.

Les aïeux, la fratrie, BDC.

Celles et ceux que j'ai eu la chance de rencontrer sur l'île de La Réunion.